"I still can't questions"

The role of Working Memory in the longitudinal development of L2 English questions in an immersion setting

Submitted by

Clare Wright,
School of Education, Communication and Language Sciences,
Newcastle University

Supervised by:
Professor Vivian Cook
Dr Martha Young-Scholten

ORIGINAL COPY TIGHTLY BOUND

Acknowledgements

My thanks go to all those without whom this PhD would not have been possible, especially my two supervisors, Professor Vivian Cook and Dr. Martha Young-Scholten, for their inspiration, insights and patient guidance. I am grateful to the Economic and Social Research Council for their financial support (grant number PTA-030-2006-00229).

I would also like to thank the many colleagues whose expertise helped at various stages of research planning and analysis, including Professor Florence Myles, and Dr. Tina Fry. I am also deeply appreciative of the support and encouragement of all my friends and, most of all, my family, Andrew, Jonathan and Alice, for their unwavering belief in me.

This thesis is dedicated to the memory of my father, Dr. Robert Conway (1931-2009).

Abstract

"I still can't questions": The role of Working Memory in the longitudinal development of L2 English questions in an immersion setting

This study adds to the growing body of research into the connection between greater working memory (WM) capacity and L2 development, by investigating correlations between WM and development of accurate wh-movement in L2 English in an immersion setting. Five WM tests were used (including Digits Back, Listening Span), with the addition of innovative story recall tasks, using both L1 and L2, to test the concept of the episodic buffer in the latest model of WM (Baddeley 2000). Three target wh-constructions were tested: short-distance and long-distance wh-questions, and subjacency violations.

Development of the target wh-constructions was tracked in a longitudinal study of thirty-two instructed Chinese speakers of English during a year's postgraduate study at universities in the UK. Participants were matched for proficiency level (IELTS 5.5 or above) and for L2 exposure in their home countries before arrival. Oral production data, timed grammaticality judgement data and WM data were collected on arrival and again after 11 months, and compared using statistical analysis.

Significant positive correlations were found between Story Recall in L1 and improvement between Time 1 and Time 2: both with improved oral question production ($r=.39$, $p<.05$) and with greater accuracy on subjacency judgements ($r=.40$, $p<.01$). However, there were no significant differences found on L2 accuracy scores between time 1 and time 2; significant improvements were only found on reaction time speeds and posthoc analysis of oral fluency. The study concludes that a year's immersion appears to favour processing existing grammatical knowledge rather than trigger acquisition of new grammatical knowledge, even for those with greater WM.

The study provides some support for the hypothesis that WM correlates with L2 grammatical development, but indicates further research is required to understand the role of WM in L2 acquisition, and the complex representation of grammatical knowledge in the L2 mind.

List of Tables and Figures

Tables in Appendices C and D

Figures

Chapter 1: Introduction

The motivation for the research studies described in this thesis is evidence of variation in second language (L2) acquisition of English wh-movement or question formation by adult Mandarin (L1) speakers.

Variability in the rate and outcome of L2 acquisition has long been identified as a defining feature of most adult L2 learners and investigated in SLA research from a wide range of perspectives (since Selinker 1972 and earlier). Unlike L1 acquisition, where the question can be asked in terms of Plato's problem of how children acquire language despite the impoverished nature of the input (Hornstein et al 2005: 2), SLA research has to ask why many second language learners, especially adults, seem unable to acquire similar linguistic expertise in that second language compared to their first, often despite much exposure to the input.

Different sub-disciplines within SLA have emerged to explain this variability, historically separating the question of what is acquired from the question of how it is acquired. Research into the nature or linguistic properties of what is being acquired by the L2 user has tended to be the focus of generative SLA research. This research may not always overlap with research into L2 development or transition from one stage to the next, which has tended to be the focus of cognitive research. Yet explaining this variability, in ways which can address both the "property" and "transition" questions of SLA (Gregg 1996), remains at the core of the SLA research agenda (R. Ellis 1994; Towell 2003; Han 2008; Piske and Young-Scholten 2009; Truscott and Sharwood Smith, in press). Key questions that remain include: what precisely is the scope of knowledge in an L2 user's mind? How do sources of linguistic knowledge interact with cognitive processes? How is the L2 acquired in real time? How does L2 acquisition overlap with L1 acquisition?

Although it is clear that there are dozens of factors affecting SLA (Spolsky 1989; Moyer 2004), core issues that have been assumed to explain systematic variability in SLA include the role of the L1, the age of the L2 learner when they are first exposed to the L2, the amount and type of exposure to input, and how the L2 is processed. So L1 transfer, age, naturalistic vs. taught exposure, and processability are seen as crucial factors to understanding variation in SLA. However, the question then arises why a particular group of L2 users, say instructed Chinese learners of English, vary in their rate and ultimate success in attainment when the group are very similar in terms of these crucial factors.

Cognitive theories of SLA assume that L2 learning is like any general skill learning, and that individuals are bound to vary because of individual differences in processing and cognitive resources such as memory. However, despite individual differences in memory and processing, there is ample evidence that there are also systematic similarities in L2 acquisition, which cognitive theories of individual differences do not fully explain. Generative theories can explain the similarities by appealing to innate rules or features of Universal Grammar, and argue that exposure to positive evidence is sufficient to trigger development. However, when exposure is very similar, and yet individuals display differences in their rate and level of L2 acquisition, generative theories cannot offer a clear explanation.

These studies presented here specifically adopt a cross-disciplinary approach to investigating causes of variability, combining a generative approach in analysing a specific formal syntactic phenomenon (wh-movement) with a psycholinguistic analysis of the role of memory, and specifically working memory, to see how far memory constraints affects variability in L2 acquisition and use in a group of instructed learners who then arrive in an immersion environment. A truly comprehensive theory of variability in SLA will need to also acknowledge (pace Spolsky 1989; Moyer 2004 amongst others) and incorporate non-linguistic social-affective factors such as socio-cultural factors, motivation, attitude, learning style and communicative strategies but these are not addressed in these studies, which focus on the interplay between implicit feature-based acquisition accounts of SLA and processing/memory accounts.

There are five assumptions driving the present research. The first assumption is that the generative programme of research describing formal parameters or features of language properties is relevant but not sufficient to explain individual variation in language acquisition, especially L2 acquisition and use in real time, and particularly for instructed L2 learners. It is claimed that much generative research into L2 acquisition is dominated by a rhetoric in which L2 acquisition can be seen as insufficient, or a failure to attain idealised L1 monolingual native norms (Cook 1997). In principle, SLA research should investigate acquisition within a framework of multicompetence (Cook 2002), in which the capacity to know more than one language not only affects that second language but may also affect the nature of the first or primary language. A key feature of the research presented here is therefore not to compare L2 learners with L1 native speakers, but to look at what may affect different L2 learners in different ways. However, it will be inevitable that some of the theoretical and empirical discussion will compare L1 and L2 users directly or indirectly.

The second assumption is that there are different types of knowledge relevant to this discussion of L2A, i.e. learned explicit declarative knowledge (or metalinguistic knowledge), and implicit competence (Sharwood Smith 1991; Schwartz 1993). I argue that L2 development, particularly for instructed learners, reveals multiple sources of knowledge utilised as a coalition of resources (Herschensohn 1999). Thus UG constraints (universal principles) operate in combination with L1 competence (which drives expectations as to how language operates grammatically) and are by default initially transferred in interpreting the L2 input. Perception and production relies on L2 interlanguage (UG constrained, implicit, language-specific), but also utilises a store of explicit learned L2 linguistic knowledge (including words, morphosyntactic chunks, morphological explanations or "grammar rules", some morphosyntactic information, e.g. past irregulars in English).

Though these complex constructs are hard to operationalise, and remain debated, one long-standing approach (following Krashen's work in the 1980s) draws a conceptual distinction between learned linguistic knowledge and implicit modular L2 competence (Schwartz 1993). This dichotomy maps to some extent with psycholinguistic constructs of explicit and implicit knowledge (Paradis 1997, 2009; N. Ellis 1994, 2005) and declarative and procedural knowledge (Ullman 2004, 2005), which are argued to operate in differing ways in an adult's stable L1 compared to a learner's emerging L2. Stable L1 and emerging L2 can be seen at opposing ends of a spectrum of proficiency factors, such as automaticity, speed, accuracy of processing and spoken fluency towards more conscious control, variability in processing or production, greater awareness and effort. As L2 users become more proficient, they could be said to resemble native speakers in the automaticity, accuracy and fluency of their language production and processing.

The present studies attempt to shed some light on these constructs by focusing on instructed L2 users who are clearly some way along the spectrum in terms of their knowledge of the L2 (adult Mandarin-speaking postgraduate students in British universities, who have typically studied English for at least six years, starting at junior high school, around the age of 12). It is not a study of initial L2 acquisition or of ultimate attainment or fossilisation, but rather focuses on the messy reality, for many language users, of a variable intermediate stage in which proficiency may still be in a transitional state.

Research investigating the learning and processing mechanisms assumed to drive growing proficiency for such learners does not yet fully account for how learners respond to exposure in different ways. Therefore, to investigate the questions of how input, acquisition and processing use memory resources in different ways in SLA in the present studies, I compare variation in acquisition of simple and complex wh-questions in English as exemplars of these dichotomies.

This research focuses on Chinese speakers of English to provide a context for examining the effect of L1 transfer and exposure. L1 transfer should occur, in accounts assuming a strong role for transfer (such as Schwartz and Sprouse's (1996) Full Transfer/Full Access theory, and Truscott and Sharwood Smith's 2004 MOGUL)[1] due to cross-linguistic formal differences between Chinese and English (typified as wh-in situ and wh-movement languages respectively). Exposure to English for such learners is predicted to produce reliance on explicit knowledge, since Chinese education in English has traditionally been instructed, heavily reliant on memorisation of rules and limited in exposure to naturalistic native English (Gu 2003).

The third assumption, following on from the above, is that Chinese speakers of English, even at advanced level, rely primarily on explicit (declarative) learned knowledge, evident through slower, more monitored, more hesitant, more variable

[1] Full Transfer/Full Access can, however, be contrasted with other theories suggesting a minimal role for L1 transfer, such as Vainikka and Young-Scholten's Organic Grammar (1996, 2005).

processing/production (Segalowitz 2003; Ullman 2005; Paradis 2009). I claim that individual differences in working memory constrain the capacity to store novel information and retrieve existing knowledge, particularly of explicit knowledge (Baddeley 2003). Working memory should thus constrain acquisition of morphosyntax within an instructed environment where the emphasis is on memorisation of explicit grammar rules (Gu 2003).

The fourth assumption focuses on how changes in input environment affect acquisition. I assume that immersion will provide added exposure, triggering L2 development of greater accuracy on target-likeness and less variability, repair or hesitation (Howard 2006). However, the role of immersion remains unclear, and whether it promotes implicit linguistic development or facilitates faster, more fluent and more efficient use of explicit knowledge is one of the questions addressed in this study.

The fifth assumption is that where previous instructed exposure is controlled by focusing on a specific L2 group thereby avoiding marked differences in age, length and type of exposure, individual variation in development after changing to an immersion environment should be affected by individual differences such as WM capacity in response to the new, and, presumably, richer input environment.

This study thus focuses on testing the claim by Miyake and Friedman (1998) that varying WM capacity is the key to variation during the course of L2A. The study is designed to bring together the interests of different research paradigms in a novel way to drive forward our understanding of the complex process that is SLA. The study also is intended to bring some clarification to the issues for the L2 users themselves, who can experience much frustration and difficulty with the target language while studying abroad. As one participant in this study concluded, distraught, in an interview at the end of a year's postgraduate study, "I still can't questions!" The aim of this study is to understand a little better why this may be so.

Chapter 2: Literature Review

2.1. Introduction

The present study draws from the wide field of research underpinning L1A and L2A in order to establish a comprehensive context for the research questions about the potential role of WM in SLA. Five assumptions were presented in the Introduction, which have been drawn from the research, and which shape the structure of this chapter. Sections 2.2 to 2.5 deal with constructs and models relating to language representation, acquisition and use in L1A, to introduce concepts, definitions and issues which, I argue, are central to SLA. Sections 2.6 to 2.11 review how these constructs are applied in SLA theories and models of mind, and provide the context for the specific research questions of this study.

Section 2.2 centres on theories of the formal properties of what is being acquired (English questions and relevant constraints and features). The section looks first at how Universal Grammar constraints are argued to operate in English questions, highlighting asymmetries between subject and object questions, matrix and embedded questions, and short and long-distance questions, which may cause difficulties for L2 learners. Then an account is presented of how these constraints are argued to operate in Chinese questions, to clarify what the significance of L1 transfer may be. I then briefly discuss alternative non-generative accounts of wh-questions, including formal and constructionist theories, to assess how effectively they might account for these asymmetries.

In section 2.3, these differing accounts of the formal properties of language are then put into context in a discussion of accounts of how the L1 is acquired in childhood, with reference to question formation where possible, comparing different models of generative and cognitive research which have been influential in SLA. In section 2.4, on language in use, four models of mind and theories of learnability are compared, which have had an impact on SLA research, including Jackendoff's (2002, 2007) Parallel Architecture model, and competition and connectionist models. I conclude that in L1, language acquisition is constrained by generative syntactic principles but that in order to understand how it comes to be acquired, particularly for complex syntax, psycholinguistic processing principles are also required.

The fifth section reviews the constructs of long-term memory and short term memory, to examine how they are understood to interface with language acquisition and use, focusing particularly on Baddeley's (2003) multicomponent model of working memory (WM). I look at how native language is stored in memory, and processed in real time, of how the constructs of implicitness and explicitness overlap with issues of automaticity or control, and how WM is argued to play a role in these different aspects of language, and why this would be relevant in SLA.

In section 2.6, I turn to theories of second language acquisition, comparing both generative and cognitive perspectives, referring back to the constructs and theories discussed in sections 2.2 to 2.5 on L1A. I evaluate differing hypotheses for their account of the role, if any, of UG, the impact of L1 transfer, explicit/implicit learning, the role of noticing and attention in SLA, automaticity, context of input for instructed or immersed learners, to show how these are relevant to my research focus on variability in L2 question formation and WM.

Section 2.7 addresses models of the mind and learnability theories for L2A, looking at the role of processability. I evaluate two models of acquisition by processing (Carroll 2001; Truscott and Sharwood Smith's MOGUL model, 2004), which are based on Jackendoff's Parallel Architecture model, and which focus on how input drives acquisition. I examine the different assumptions about WM in these models.

In section 2.8, the issues raised in sections 2.6 and 2.7 are integrated in a discussion of two models that assume a bi-modal construct of explicit and implicit linguistic knowledge (Schwartz 1993; Ullman 2001, 2005). These models are compared to Truscott and Sharwood Smith's MOGUL model and the hypothesis is presented that L2 acquisition depends on a "coalition of resources" (Herschensohn 1999) in a multicompetent mind (Cook 2002) and is constrained by processing (Truscott and Sharwood Smith 2004). The hypothesis assumes that both formal and processing constraints are crucial to understanding variability in L2 acquisition, and underpins the focus of this study of how memory, and working memory specifically, can play a role in the process of second language acquisition.

The remaining sections, 2.9 and 2.10, turn to justify why WM is assumed to play a critical role in L2 acquisition and discusses selected empirical studies which have investigated the role of WM in L2. Finally, in section 2.11, the specific research questions of my study are presented.

2.2. Theories of linguistic representation

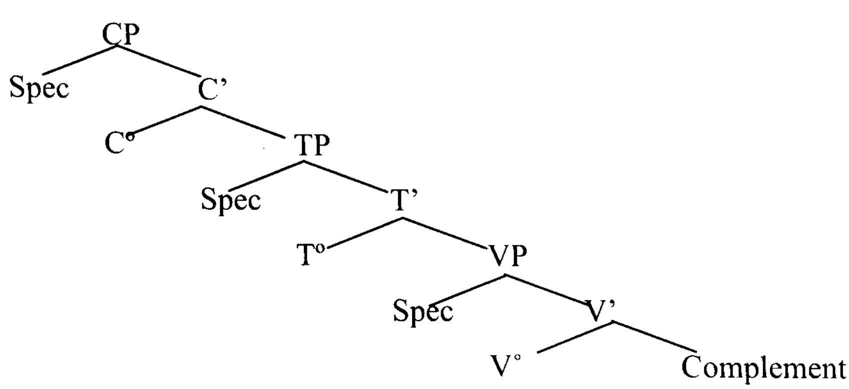
English and Chinese are taken as examples of opposing types of wh-question formation – English is known as a wh-movement language, whereas Chinese is known as wh-in situ. In this section, I will first review generative theoretical accounts of English question formation, and constraints on wh-movement, and then turn to accounts of Chinese question formation to show the comparison.

2.2.1. Generative syntactic accounts of question formation in English

Generative approaches to language acquisition argue that human language is constrained by an innate system of Universal Grammar (UG) – a linguistic domain-specific module in the mind (Fodor 1983) which contains universal combinatorial abstract principles. These principles constrain how sentences are constructed, and why surface word order may not always logically represent the underlying syntactic structure, using implicit unconscious features and processes.

Generative theory argues that syntactic structures are binary-branching tree structures, with different phrasal projections: verb phrase (VP), finite tense phrase (TP) and complementiser phrase (CP), as shown in (1) below.

(1)



These projections are constrained by universal principles, constraints or features which may apply at different stages of the projection.

Chinese and English are standardly seen as parametrically opposed in question construction in that English overtly fronts wh-expressions and modals/auxiliaries which are moved from their underlying position in declarative sentence position, whereas Chinese uses question particles and/or leaves the wh-expressions in situ. Universal principles are also seen as constraining what expressions can move where, especially in long-distance wh-movement (e.g *What did John say Mary liked?),* as will be explained in more detail below.

In English questions, the overt word order in questions is different from the underlying word order, and this is accounted for by a theory of movement (originally Move α, Chomsky 1981). Current mainstream syntactic theories of movement based on Chomsky's Minimalist Program (1995) assume that grammatical expressions or derivations are created by a computational system with two principal operations, Merge and Move. Grammatical expressions are generated by this computational system operating on items in the lexicon to pair up form and meaning through two systems (Hornstein et al 2006: 15). These are the Conceptual-Intentional system (Logical Form or LF) and the Articulatory-Perceptual system (Phonetic Form or PF). Constraints can apply either at LF or PF and explain why syntactic structure (LF) can be different to what is pronounced (PF). Lexical items are assumed to be comprised of arrays of phonological, semantic and syntactic features, which are stored in the lexicon with these syntactic features already in place; these features need to be "checked" against feature licensing requirements in the governing head or specifier (Hornstein et al 2006: 29). Features can be strong [+] or weak [-]: strong features are argued to trigger overt movement seen at PF; weak features, however, can be seen by constraints on grammatical formulations at LF. Constructions are typed as declarative or interrogative within the C domain (Cheng 1997; Platzack 2001). A strong interrogative [Q] feature, located within the C domain, is typically argued to drive overt wh-movement as seen in English. However, the C domain is believed to be "vulnerable", with variability found for children with SLI and adult L2 learners (Platzack 2001), and so how CP is structured and projected may require more subtle analysis, as I show below.

English has in general two types of questions: yes-no questions and wh-questions (as well as tag questions, and echo questions using intonation only – these last two do not involve movement and are not discussed here). Thus a statement such as (2) below can generate questions such as (3) and (4).

(2) John has eaten cake.

(3) Has John eaten cake?

(4) What has John eaten?

2.2.2. Yes-no questions

This type of question, seen in (3) shows overt subject/auxiliary movement, also known as T to C movement (Adger 2003: 294-6). The question or interrogative function is argued to derive from a null question particle Q located in the head C, which has a strong Tense [+tns] feature (Adger 2003: 295; Radford 2004: 153). The strong T feature requires the head T constituent of TP to move from T to C (head movement).[2]

Where T is overtly filled with an auxiliary, this causes no problem, as seen in (5).

(5) Has John eaten cake?

But consider (6) and (7):

(6) John eats cake.

(7) Does John eat cake?

In English declaratives, unlike in questions, the Tense feature is seen as weak [-tns], and therefore, in declaratives without auxiliaries, there is no overt element showing Tense. In standard accounts of English, Tense is seen as an affix, i.e. a grammatical morpheme which cannot stand on its own (Radford 2004: 433) and must attach to an overt verbal element. Therefore, if there is no auxiliary, the morpheme marking T, here showing 3rd person singular present (-*s*), lowers or "hops" down to affix to the verb in the VP. However, affix hopping must be local, so the Tense affix cannot hop all the

---

[2] Standard syntactic accounts now split VP into vP and VP (Adger 2003: 133), or refer to verbal inflectional tense phrases as IP, but I use VP and TP throughout for consistency with most of the studies referred to here, since the differentiation is not, I assume, relevant to my study.

way from VP to C. Therefore, a dummy auxiliary "do" is called to support the tense affix, as seen in (7) above.[3]

2.2.3. Wh-questions

For wh-movement, there is argued to be a universal requirement for wh-expressions to move to the specifier of CP from their underlying trace position (Chomsky 1981, 1986, 1995; Rizzi 1990, 1996; Pesetsky and Torrego 2001, Adger 2003).[4] In English the wh-expression is assumed to have a strong [+wh] feature, thus triggering overt movement of the relevant wh-phrase from its underlying position (e.g. object of VP) to the specifier of CP (SpecCP). Once the features on the trace wh-expression are matched or "checked off" against the features in the landing site, the original trace or copy is deleted. The strong [Q] feature also requires T to C movement, generating subject-auxiliary inversion or do-support, as outlined above for yes-no questions, and shown below in (8) and (9).

(8) What does John eat *t*?

(9) Who has Mary seen *t*?

However, there is a potential asymmetry between subject and object questions. Subject questions in main clauses are argued to be generated slightly differently (Pesetsky and Torrego 2001). The subject pronoun *who* carries a tense feature, because subjects must always appear in a tensed clause (Radford 2004: 211). Therefore since the pronoun can check both [wh] and [tns] features, no do-support is required, and the tense affix is free to affix-hop down to the main verb, as shown in (10) below.

(10) Who *t* gives Mary coffee?

---

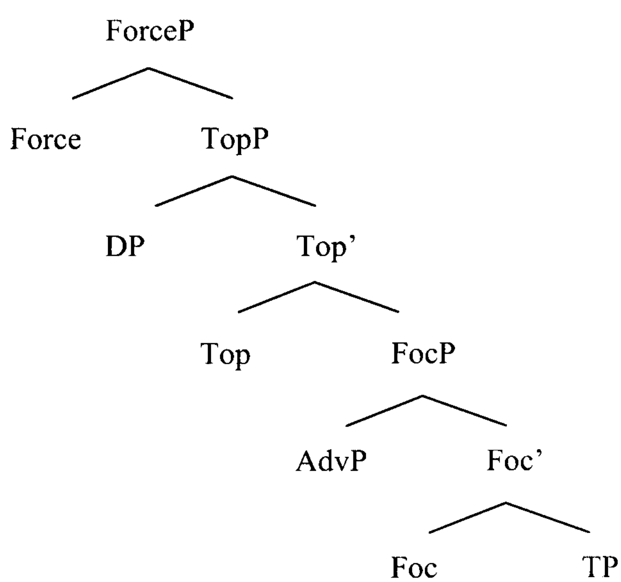
[3] The overt use of "do" was more widely prevalent as an auxiliary in older variants of English, and can still be found in some dialects and creoles (Radford 2004: 186) but is assumed here not to be relevant to L2A.

[4] There is ongoing debate over the precise nature of how interrogatives are formed (see Jayaseelan 2008 for an alternative account of English questions in which question meaning is signalled using a universal question particle which is null in English). There is also ongoing discussion of why movement in English should be overt (Hornstein et al 2006: 293-298; Radford 2004: 206). However, in the context of this study, I adopt the standard explanation that, in a formal generative theory of wh-movement in English, movement is triggered by a strong wh-feature.

The analysis thus far has assumed a single CP structure. However, recent work by Rizzi (1996, 2001) and others suggests that CP needs to be redefined into multiple phrase structures. Rizzi's "Split CP Hypothesis" (1996) identifies that complementisers play different roles (or show different force) and CP should be split into ForceP (to show if the clause is declarative, interrogative, imperative or exclamative), containing TopicP (showing topicalised elements), then FocusP (such as fronted adverbial phrases which trigger auxiliary raising through a strong Tense feature).

The full tree diagram for the split CP is shown in (11), with sentence examples in (12) and (13) below (adapted from Radford 2004: 328-330):

(11)



(12) That behaviour she never again tolerated *t*. (Topic fronting)

(13) Never again did she *t* tolerate that behaviour
(Focus fronting and auxiliary movement)

Rizzi (1996: 299) identifies that wh-elements are moved to fill the specifier of the lowest element, FocusP, given the ungrammaticality of (14), (15), (16) and the grammaticality of (17):

(14) *Who that behaviour *t* never again tolerated?
(*wh-movement over Topic fronting)

(15) *What never again did she tolerate *t*?

(*double Focus fronting)

(16) *Can that kind of behaviour we tolerate *t* in a civilised society?

(Focus before topic)

(17) What did she never again tolerate *t*?

(wh-movement blocking adverbial fronting)

Additional evidence that a moved wh-element can only be found on FocusP can be adduced from the ungrammaticality of multiple wh-questions in English which take double complements, supposing that a nominal wh-element could be topicalised while another wh-element licitly occupies Specifier of FocusP, such as in (18) and (19) below:

(18) *What where does she put *t t*? (She puts the coffee on the table)

(19) *What who does she give *t t*? (She gives Mary the coffee)

However, the syntactic arguments surrounding multiple wh-questions are highly complex (see e.g. Aoun and Li 2003; Boeckx 2003), and largely fall outside the scope of this study, so I go no further in the theoretical account of these structures. The point is to stress that in English, TopicP and FocusP operate with different constraints on movement, which clarifies what can be fronted via wh-movement as opposed to topicalisation. However, as we will see below, in Chinese, topicalisation is a much more prominent syntactic strategy than in English, and constraints on wh-elements are different; these cross-linguistic differences could be important in a discussion of acquisition of English features triggering wh-movement by Chinese speakers, and the potential role of L1 transfer (argued by Hawkins and Chan 1997, and explained in more detail in section 5.4).

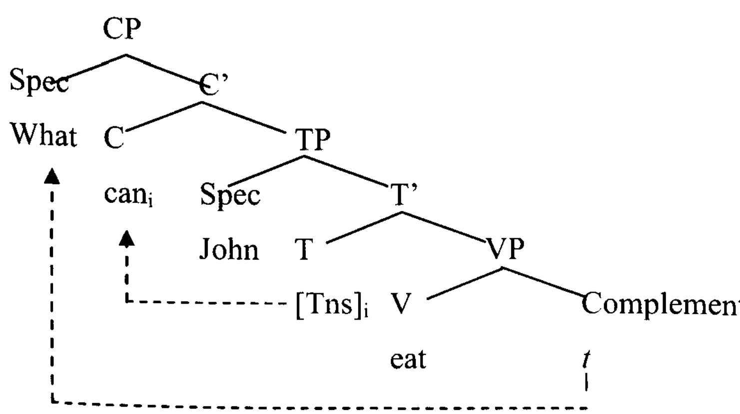
In light of Rizzi's split CP hypothesis, it could be suggested that, strictly speaking, the Wh-Criterion's definition that the wh-element is in SpecCP could be refined as relating to Specifier of FocusP. However, in most accounts of formal syntax, it remains usual to refer to wh-movement into SpecCP (Radford 2004, Aoun and Li 2003), or perhaps to "the clause-periphery" (Huang, Li and Li 2009), and for ease of reading, I retain the standard terms where relevant of CP and SpecCP.

2.2.4. Embedded /long-distance questions

The types of questions shown so far are known as short-distance questions or short movement, with only one CP structure, involving a single instance of movement and verb raising from T to C generating do-support, as in (20), and illustrated in the tree in (21) below.

(20) What can John eat *t*?

(21)



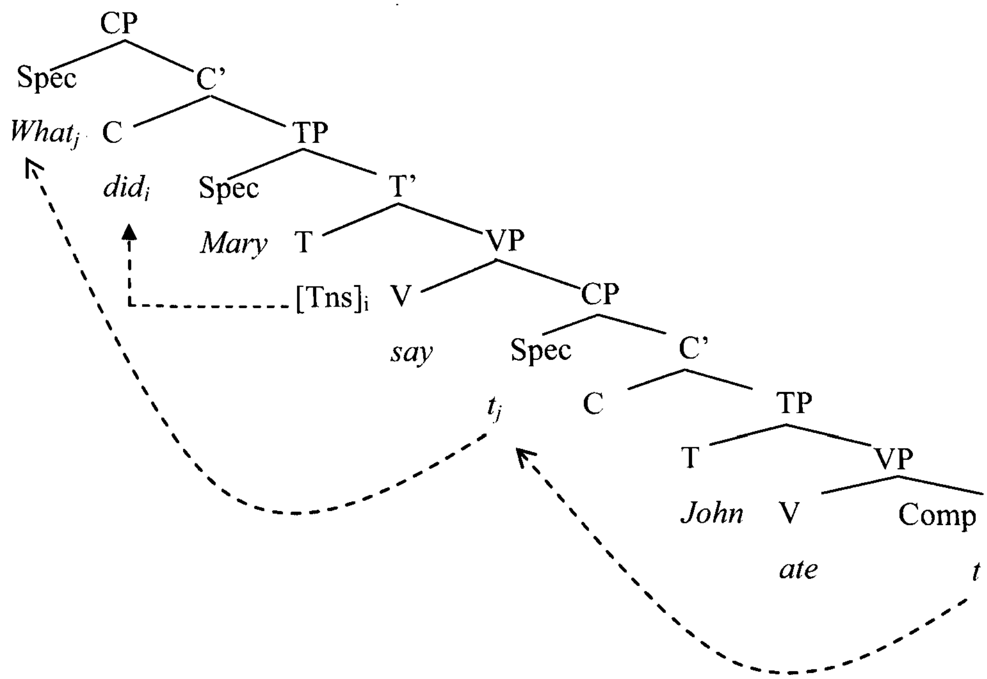
How do the features and constraints described above apply in embedded or long-distance questions such as (22), which have more than one CP?

(22) What$_j$ did Mary say John ate t_j?

Here we see two CP clauses – the main clause "What did Mary say.." and the embedded clause ".. John ate *t*". Rather than explaining this as a single movement of the wh-element directly from the end to the front of the projection in a single jump, it is argued, through the concept of "successive cycling" (Haegeman 1994; Radford 2004; Hornstein et al 2006: 358) that movement can only take place up to the nearest equivalent position in the projection. Thus the trace or copy of the original wh-element "what" moves first to the intermediate SpecCP above "John ate *t*" and then moves again to the SpecCP of the main clause as shown in (23) below and illustrated in (23) below

(23) What$_j$ did Mary say t_j John ate t?

(24)



Evidence of the successive cyclic hypothesis of long-distance wh-movement, with copies or traces in intermediate position, can be found in in the phenomenon of wh-copying (Radford 2004: 397-98). A number of languages, including Afrikaans, leave the moved copy of a wh-pronoun overt in the intermediate position, as shown in (25) below, taken from Radford (2004):

(25) *Waarvoor* dink julle *waarvoor* werk ons?
What-for think you what-for work we?
"What do you think we are working for?" (Afrikaans)

A similar phenomenon has been noted in child acquisition (Thornton 1995: 147) where children have been noted as producing wh-copies such as (26) and (27):

(26) *What* do you think *what* Cookie Monster eats?
(27) *How* do you think *how* Superman fixed the car?

A notable asymmetry exists between matrix questions and embedded questions; in embedded questions wh-movement and head movement do not appear to operate in the same way. Hence, in standard English, we see that head movement in questions only applies to the matrix verb: i.e. we do not standardly find (28):

(28) *What did Mary say did John eat?

Most explanations suggest that this is due to the syntactic function of the embedded question as complement of the matrix verb, since the strong T feature of Q in C is only found in main clauses, not complements (Radford 2004: 199), although it is not clear why this should be so. Radford (1997: 287-291) proposes that there is a null complementiser Ø in the head of CP; this null complementiser, Ø, adjoins in C to the Q affix in complement clauses through merger (ibid: 284), which therefore blocks T-to-C movement.

2.2.5. Constraints on wh-movement

Wh-elements can thus be shown to move from their original position through successive SpecCPs. However, it has long been observed that not all wh-elements can move in the same way; movement is constrained as to which wh-elements can be moved where. Compare the sentences below:

(29) What does John believe Mary likes *t*?

(30) *Who does John believe the claim that Mary likes *t*?

(31) *What was a dish of t cooked by Ann?

Ross (1967) observed the ungrammaticality of extraction from complex NPs and embedded wh-clauses such as indirect questions and relative clauses (known as "wh-islands") compared to grammatical long-distance extraction. Chomsky (1986) proposed the principle of Subjacency, requiring the wh-extraction to land in the specifier of each CP crossed, leaving a trace in each landing site, without crossing more than one IP or NP node at once. Huang (1982) developed this into a general Constraint on Extraction Domains that argued that only complements allow material to be extracted, not specifiers or adjuncts.

Adjunct extraction from wh-islands and adjunct islands (clauses introduced by subordinating conjuctions such as *because*, and certain overt complementisers such as *if*) are seen as strongly ungrammatical, as shown below:

(32) *What did John see [$_{NP}$ the girl who cooked *t*]?

(33) *What did Mary go home because [TP she needed to do *t*]?

(34) *What did Tom wonder if [TP Mary bought *t*]?

Extraction from complex NP islands is also strongly ungrammatical, either through extraction from object position (35) or subject position (36), as shown below:

(35) *Who does John believe [NP the claim that Mary likes *t*]?

(36) *What was [NP a dish of *i* cooked by Ann]?

However, there are asymmetries within these island constraints. Movement over the overt complementiser *that* is seen as more weakly ungrammatical, and only constrains movement on subjects, as shown below.

(37) Who do you believe *t* saw Tom? (subject)

(38) *Who do you believe that *t* saw Tom? (subject)

(39) Who do you believe John saw *t*? (object)

(40) Who do you believe that Jim saw *t*? (object)

Similarly, extraction of objects from complex NPs are seen as weaker violations than subject extraction, as shown below:

(41) ?What did John like the book about *t*? (object)

These asymmetries within the Constraint on Extraction are traditionally explained by the Empty Category Principle (Chomsky 1981), which places constraints in differing ways on movement of subjects and objects. Traces must be properly governed, either lexically (as a complement) or by antecedent (bound by a category which governs the trace). Object traces are lexically governed as arguments of the verb; subject traces are not properly governed if not bound by their antecedent. The same rule applies to adjuncts, which must also be antecedent-governed. Hence subject and adjunct extraction produces stronger subjacency violations than object extraction (Huang 1995: 152).

Huang's (1982) Constraint on Extraction Domain remains at the heart of the most recent reanalyses of wh-movement within the Minimalist Program, where, as outlined above, a strong wh-feature on interrogative wh-elements in English drives overt movement seen in its final form at PF. In comparison, Chinese has a weak wh-feature, and thus does not trigger overt movement.[5]

In generative theories of SLA, the task of acquisition for Chinese learners of English is to acquire these cross-linguistic feature differences, especially in identifying the asymmetries between subject and object questions, matrix and embedded questions, and constraints on movement. Subjacency constraints were believed to be language-specific to languages with overt movement – i.e. English was constrained by Subjacency, whereas Chinese was not. However, Huang (1982, 1995) argued that the general Constraint on Extraction applied in Chinese, as did Subjacency, but at LF not PF. I turn now to an overview of wh-structures in Chinese to explore the context of this debate in more detail, and to draw attention to cross-linguistic differences that could cause difficulty for Chinese learners of L2 English, through L1 transfer.

2.2.6. Generative syntactic accounts in L1 Chinese

Chinese and English have been assumed to show cross-linguistic parametric, language-specific, differences in wh-questions. In Chinese, questions have traditionally been seen as remaining in-situ at sentence level (Haegeman 1994: 447; Radford 2004: 198). There are four different types of question forms in Chinese: closed questions with a question particle, alternative or disjunctive questions, verb-not-verb constructions and wh-in-situ.

[5] Some doubt remains as to whether, theoretically, explaining movement in terms of features adds any greater understanding to what actually happens, as "we have no account in terms of feature strength as to why… some features are strong in English and weak in Chinese" (Hornstein et al 2006: 42).

2.2.7. Closed questions

Closed yes-no questions are formed by inserting the question particle "ma" at the end of the clause but leaving declarative word order overtly unchanged, as shown below:

(42) ta zhu zher.
He live here
"He lives here." (Huang et al 2009: 238)

(43) ta zhu zher ma?
He live here Q
"Does he live here?" (Huang et al 2009: 238)

2.2.8. Alternative (disjunctive) questions

Chinese disjunctive questions are formed by linking two or more constituents with *haishi* ("or"). The same structure is used for declaratives or questions, with context and/or intonation determining if it is a question. Constituents can consist of S, VP, PP, NP or V (ibid: 243), as illustrated below:

(44) Zhangsan zai jiali shuijiao haishi zai gongsi shangban?
Zhangsan at home sleep or at firm work
"Is Zhangsan sleeping at home or working at the firm?"

(45) Zhangsan xihuan haishi taoyan Lisi?
Zhangsan likes or detests Lisi
"Does Zhangsan like or detest Lisi?"

More restricted disjunctive questions may also be formed without the conjunction *haishi* as seen below:

(46) Ni jintian chi fan chi mian?
You today eat rice eat noodles
"Would you like to eat rice or noodles today?"

(47) Ni xihuan Zhangsan taoyan Zhangsan?

You like Zhangsan detest Zhangsan

"Do you like or detest Zhangsan?"

These types of disjunctive questions without *haishi* are known as "juxtaposed choice questions" (ibid) and are constrained by needing some degree of phonological similarity, and require either both verbs or objects to be the same, as seen above. If either the verb or the object is different, the result is ungrammatical:

(48) *ni xihuan Zhangsan taoyan Lisi?

You like Zhangsan detest Lisi?

Intended reading – "Do you like Zhangsan or detest Lisi?" (ibid: 243)

2.2.9. Verb-not verb questions

The discourse intention of indicating choice through disjunctive questions can also be expressed in specific "verb-not verb" (A-not A) construction, using the negative particle *bu*, with fixed word order, illustrated below:

(49) Zhangsan mai bu mai shu?

Zhangsan buy not buy book

"Does Zhangsan buy or not buy books?" (ibid: 244)

The conjunction *haishi* can also be found with verb-not verb questions; however, word order is not fixed:

(50) Zhangsan bu mai haishi mai shu?

Zhangsan not buy or buy book

"Does Zhangsan buy or not buy books?" (ibid)

It is argued that the two types differ in terms of island constraints as well as their ability to reorder their constituents (ibid: 248). *Haishi* questions are free from island constraints and retain a range of direct or indirect interpretations whether used in matrix or embedded questions. True A-not A questions can only be interpreted as indirect questions in embedded contexts such as sentential subjects or relative clauses (ibid: 246), as shown below:

(51) Zhangsan bu xiaode [ni lai haishi bu lai
Zhangsan not know [you come or not come]
"Zhangsan doesn't know whether you will come or not.

(52) *[ta lai bu lai] bijiao hao?
he come not come more good?
=?Intended: "Is it better that s/he comes or that s/he doesn't?"

(53) *ni bijiao xihuan [lai bu lai de nei-ge ren]?
you more like come not come ASP that-CL person
=?Intended: "Do you prefer the person that will come or the one who will not"?

There are further analyses of disjunctive questions, in which verbal elements can be ellipted, and negative particles used on their own, but these phenomena do not appear to play a major role in movement constraints in Chinese; neither, to my knowledge, have they been implicated in possible transfer effects in SLA, so I do not go into detail here.

2.2.10. Wh-in situ

Chinese wh-questions are formed, unlike in English, by leaving the interrogative wh-expression overtly in situ at PF. The strong [+Q] feature present in the head of C in English selectively binding most wh-elements is argued to trigger movement. In Chinese the Q expression is argued to be unspecified (Qu) or "unselective" and so does not trigger movement (Huang et al 2009: 278). In addition, there is no strong T feature triggering T to C movement. Thus in Chinese, wh-in situ questions do not show the asymmetries in English between subject and object questions, or between matrix and embedded questions as shown below:

(54) Zhangsan kanjian-le shei?
Zhangsan see-ASP who?
"Who did Zhangsan see?" (Huang et al 2009: 261)

(55)	ni	xiangxin	ta	hui	shuo	shenme?
	You	think	he	will	say	what?

"What do you think he will say?" (Radford, 1997: 18)

(56)	Wo	xiang-zhidao	Lisi	mai-le	shenme.
	I	wonder	Lisi	buy-ASP	what?

I wonder what Lisi bought. (Crain and Lillo-Martin, 1999:204)

2.2.11. Covert constraints on movement

Huang (1982, 1995; Huang et al 2009) has argued that Chinese shows covert wh-movement at LF. The justification for LF movement came from observed syntactic-semantic restrictions seen in both Chinese and English on the complements of certain verbs, such as *think*, *wonder* and *remember*. *Think* cannot have a wh-element as a complement, whereas *wonder* must be followed by a wh-element, and *remember* can have either, as shown below (taken from Huang et al 2009):

(57) a. What does John think Mary bought?
b. *Does John think what Mary bought?

(58) a. John wonders what Mary bought.
b. *What does John wonder Mary bought?

(59) a. What does John remember Mary bought?
b. Does John remember what Mary bought?

The equivalents in Chinese all have the same surface word order, as shown below:

(60)	Zhangsan	yiwei	Lisi	mai-le	shenme
	Zhangsan	think	Lisi	buy-ASP	what

"What does Zhangsan think Lisi bought?"

(61) Zhangsan xiang-zhidao Lisi mai-le shenme

Zhangsan wonder Lisi buy-ASP what

"Zhangsan wonders what Lisi bought."

(62) Zhangsan jide Lisi mai-le shenme

Zhangsan remember Lisi buy-ASP what

"Zhangsan remembers what Lisi bought."

"What does Zhangsan remember Lisi bought?" (Huang et al 2009: 262)

However, the discourse interpretations are distinct, in that (60) must be interpreted as a direct question, (61) must be interpreted as an embedded question, and (62) may be interpreted either way. Additional evidence that the semantically distinct interpretations reflect covert LF movement comes from restrictions on the scope of adjunct wh-elements, which reflect the constraint on overt extraction of adjuncts, seen in English. In Chinese, an adjunct phrase such as *weishenme* ("why") cannot be used to form a direct question about the adjunct phrase when it occurs within a syntactic island such as a relative clause (Huang et al 2009: 264), as shown below:

(63) *ni zui xihuan [weishenme mai shu de ren]?

you most like why buy book ASP person

"Why do you like [the person who bought the books]?

However, a number of problems remain with Huang's argument that wh-extraction constraints are equivalent in Chinese and English, differing only in whether the constraint is revealed overtly at PF or covertly at LF (see overview by Cheng 2009). Aoun and Li (1993) maintain that wh-operators in Chinese do not move at all but remain wholly in situ. They suggest that the similarity of constraints on adjunct extraction between movement and in-situ languages is overweighed by the difference found on argument extraction.

Another interpretation turns to other principles of movement such as quantifier raising, which show differing scope of restrictions to wh-movement (Huang et al 2009: 270-272). Wh-argument elements can be both interrogative and existential quantifiers (ibid: 271). Thus movement is allowed by Quantifier Raising, and does not fall under the constraints of Subjacency. Adjunct elements could be seen as quantifier elements, and would similarly not be constrained by subjacency; however, they are seen as not licensed by proper antecedent government (violating the Empty Category Principle). Thus the observed ungrammaticality of adjunct extraction derives from the requirements of ECP, not subjacency. Relative clauses cannot be seen as quantifier elements, and thus display "a full range of island constraints" (ibid: 273). This accounts for why most wh-expressions in situ do not show any extraction constraints. However, as shown above, adjunct extraction of e.g. *weishenme* ("why") is not allowed. Huang et al (2009) review arguments that abstract movement strategies must apply to account for these observed locality effects (ibid: 281) either because adjuncts are in fact specified to be interrogatives, and are therefore constrained by general principles on movement, or because wh-adjuncts are bound by an abstract Qu operator, and it is the Qu operator that moves.

2.2.12. Crosslinguistic differences on Tense and Topicalisation

As we have argued above, generative theory holds that there are crosslinguistic differences arising from different feature specifications on question formation, wh-movement, constraints on interpretation, and word order between Chinese and English. In addition, there are differences in showing Tense, as mentioned above. Chinese is underspecified for Tense, but shows Aspect, using a system of aspect markers which operate pre-or post-verbally (Huang et al 2009: 101-5). These markers have traditionally been argued to show affix-lowering, similar to English in declarative sentences (Huang et al 2009: 102). Unlike English questions, where T-to-C raising occurs, Chinese aspect markers remain unaffected in interrogatives, as can be seen in the comparison of the two following examples below (adapted from Huang et al: 101).

(64) a. ta zai chang ge
he at sing song
"he was singing"

b. ta zai chang ge ma?
he at sing song Q
"was he singing?"

Similarly, modals used between subject and main verb are similar to English, but some modals can be found in sentence final or sentence-initial position, which is not expected if located in T. Thus there appear to be two groups of modals, where the first type behaves as lexical verbs, taking a clause as subject (and can be known as "raising modals": ibid 109), such as *gai* ("should") and *hui* ("be likely"). These modals can also be found sentence-initially (ibid: 108) but only for question interpretation if the following clause is understood as the object of the modal as illustrated below (Huang et al 2009: 107-108).

(65) zhe-ge ren gai-bu-gai shou fa?
this-CL person should-not-should receive punishment
"should this person receive the punishment?"

(66) hui-bu-hui ta xiang chuguo?
be.likely-not-be.likely he want go.abroad
"is it likely that he wants to go abroad?"

A second smaller group, acting as "control" modals, have specific semantic requirements to do with sentience and free will, including *yao* ("wish") or *neng* ("be able to") (ibid: 110), with tight reference to subject NP. These can only occur in canonical position between subject and verb, using intonation to determine question context, and sentence-initial position is ungrammatical, as illustrated below.

(67) wo neng chang yi-shou xiaoqu
I can sing one-CL little.song
"I can sing a little song"

(68) *neng-bu-neng ta chang yi-shou xiaoqu?"
can-not-can he sing one-CL little.song?
Intended: " Can he sing a little song?"

The implications of this semantically-driven subdivision of modals may be important in explaining possible transfer effects on word ordering in L2 English questions using similar words (such as *can* or *might*) if non-raising of modals is accepted in conjunction with wh-fronting. Evidence that other non-sentient type modals and auxiliaries marking tense or aspect (e.g. *have/will/be*) are raised without difficulty in comparison would argue that this semantic contrast plays a role. Evidence that all modals or auxiliaries are used similarly (either raised or not raised) would suggest that the semantic contrast in Chinese does not play a major role.

There may also be transfer effects from Chinese topicalisation which have been argued to apply in L2A (Hawkins and Chan 1997), where object fronting through topicalisation may be relevant and helpful as an alternative processing strategy if full wh-movement features are not yet in place. This could produce ungrammatical questions in L2 English where the wh-expression is fronted, and the verbal expression is marked for tense, but in lowered position, as in declarative sentences (perhaps as a default position), and there is no do-support, illustrated below.

(69) a. *What Mary likes?
b. *What Tom can eat?
c. *What John said Tom ate?

These "in situ" formations would show a different (perhaps later) stage of development from a bare VP or noninflected verbal expression, such as shown below.

(70) *What John say Tom eat?

Evidence of tense-marked but non-raised verbs in questions in conjunction with awareness of English constraints on movement (showing acquisition of strong [+wh] could be argued to show that strong [+T] features for T to C movement are acquired differently than the strong [+wh] feature. It is not discussed in the theoretical generative literature, as far as I have yet been able to establish, whether these features are in any way seen as separable. However, there is some discussion in Lardiere (2009) on the notion of feature reassembly which relates to this suggestion, especially for L2 acquisition of other forms of overt morphology, in which bundles of features typically associated in a single cluster (e.g. for agreement, tense, number) are separated into the separate feature settings, and acquired separately.

An additional grammatical transfer effect may be found from the use of existential verbs (*you* – "there be") for clause initial subjects (Yip 1995; R. Hawkins 2001). This is illustrated below, which would be ungrammatical without *you*.

(71)	You	yi-ge	ren	lai	le.
	There.be	one-CL	person	come	ASP
	"Someone has come."				

In L2 English use of copula questions (a frequent strategy as observed above, and below, section 6), there may be an effect of L1 transfer combining with learned copula chunk questions (such as *Is there ..?*) used to mark a subject of a question, rather than verb-raising as required by English [+Q] features.

2.2.13. Implications of English/Chinese differences between for generative SLA

Some of the crosslinguistic differences outlined above are likely to be clear for instructed learners of English from very early on, such as fronting of wh-question words, and subject-auxiliary inversion in yes-no questions. In order to acquire English features for wh-movement, a strong [+wh] feature on the wh-expression needs to be acquired, and the entailed strong [+Q] feature setting is also required to drive verb raising or T to C movement.

National curriculum requirements for Chinese and Taiwanese school students show that wh-questions, both short and long-distance, are introduced in the first two semesters of junior high school, starting with copula questions (e.g. *What's your name*?) and then lexical questions (e.g. *What do you like*?) and these forms are part of the national examinations at the end of junior high school (around 13 years of age, after about 3 years of schooling) (Chinese Ministry of Education 2003; see also Nani 2006 - Junior High School English School Books 1 and 2). To some extent the presentation, as seen for example in Taiwanese high school text books (Nani 2006), could be seen to promote chunk learning, as certain key forms are repeatedly used with a few copula and frequent lexical verbs. Disentangling correctly memorised chunk production from creative construction using innate features may be difficult in analysing the L2 data. In addition, ungrammatical subjacency-constrained questions would, by definition, not form part of positive evidence from primary linguistic data, although it is possible that attempted constructions that violate subjacency constraints may be explained as ungrammatical. I have not been able to obtain examples in the literature or real-time observations of any instances of such metalinguistic explanations of subjacency violations. I will assume for the purposes of this study (following Laurence and Margolis 2001; White 2003) that the input received by instructed Chinese learners will be underdetermined for constructions that show evidence of how subjacency constraints operate in English. Learners may be able to rely on taught knowledge for short-distance questions, but it is argued that successful acquisition of all the phenomena discussed above will show that UG constraints on acquisition are a necessary element within L2 acquisition.

Therefore, in concluding this section, it is clear that for Chinese learners of English, generative theories of acquisition must answer four main questions. Firstly do learners acquire strong [+Q] and strong [+wh] features found in English, and do learners show sensitivity to constraints on movement, in particular subjacency constraints on extractions out of wh-islands, complex NPs and adjuncts?

Secondly, might they show a difference between acquisition of [+wh] features constraining movement, especially long-distance movement, compared to the strong [+Q] features which entail [+T] driving T to C movement?

Thirdly, are learners successful in acquiring the asymmetries found in English between subject and object questions, and between matrix and embedded questions?

Fourthly, how far might transfer of L1 phenomena (disambiguation between disjunctive and interrogative particles, topicalisation, underspecification of T, argument/adjunct asymmetry on long-distance interpretations, semantic-syntactic constraints on modal movement) affect acquisition of target-like English wh-movement?

However, there are other accounts of wh-movement, outside the generative framework, particularly for subjacency constraints, which are briefly discussed below to identify if they may have different predictions for Chinese learners of English from those presented above.

2.3. Alternative theories of wh-movement

Some theories, such as Culicover and Jackendoff's Simpler Syntax Hypothesis (2005), Pollard and Sag's Head-driven Phrase Structure Grammar (1994), and Lexical Functional Grammar (Bresnan 1982, 2001) have attempted to draw up alternative formal systems of grammar, which are still based on the notion of symbolic structures, but have created a different syntactic paradigm to the Chomskyan generative paradigm. These theories are more closely derived from surface word order and reflect an interest in theoretical explanations that "lead to a more direct relation between the theory of linguistic structure and a theory of processing" (Culicover and Jackendoff 2005: 307). Other theories discount any notion of a language module, and focus more on language learning as a general cognitive skill, limited by processing constraints, in which underlying constructions are shaped by frequency and salience in the input, or other constraints, such as pragmatic understanding (e.g. Tomasello 2003; O'Grady 2005). These lines of research derive from early work on language functions and typologies by Greenberg (1963) and others (see also, e.g., Halliday 1985; Langacker 1987; Fillmore, Kay and O'Connor 1988; Newmeyer 1998).

2.3.1. Non-generative formal accounts

Culicover and Jackendoff (2005) propose the Simpler Syntax Hypothesis, replacing the notion of a hierarchical tree with a "flat" structure (ibid: 110), in which they "give up entirely the notion of movement in syntax, and with it the notion of any 'hidden levels' of syntactic structure." (ibid: 111). They retain single-tier treelets, or nodes reflecting lexical categories dominated by a phrase category (X and XP) but there is no intermediate X' tier (ibid: 110). This model (and Jackendoff's associated Parallel Architecture Model) has been influential in some recent models of SLA, and so is discussed in some detail here, and again in section 2.5 below.

Rather than a bimodal model of LF and PF checking features to generate movement on items stored in the lexicon, the Simpler Syntax Hypothesis (SSH) suggests a trimodal model of equal but separate submodules: the conceptual system (CS), syntax (SS) and phonology (PS), linked by interfaces or "correspondence rules" primarily linking form and meaning. SSH reformulates the explanation of question word order as determined by the structure of questions in CS (ibid: 309). Constraints on wh-structures are derived from two key principles – indirect licensing and traces. Indirect licensing is proposed to explain the apparent incompleteness of fronted or "orphan" phrases (ibid: 257) that need reference to an antecedent elsewhere not only for their semantic interpretation, but also to their "syntactic well-formedness." (ibid: 235). Traces are retained in their hypothesis but only as the target of a single linear "syntax–semantics correspondence rule" (ibid: 301), so that there are no intermediate traces (ibid: 302).

SSH does not shed much light on the asymmetries identified earlier in matrix questions, and overtly avoids discussion of certain aspects (the "exact formulation of inversion" is left aside: ibid: 309). However constraints on long-distance movement are still held to apply (such as the Complex NP island constraint), although are explained here without recourse to the notion of movement. Instead SSH adopts the "slash (/) category" put forward by Pollard and Sag and others within the framework of Head-driven Phrase Structure Grammar (see below). Slash-categories act as a linking chain of a slash feature (XP/ t_{YP}) set on a trace YP up through the constituent XP phrases dominating the original trace until satisfied or "saturated" by the wh-fronted element. Constraints on ungrammatical wh-fronting, such as in complex NPs (e.g. *"What did the book about t please Mary?") reflect limitations on the transfer of a slash across "a particular configuration" (ibid: 332).

Although merely substituting the notion of "constraint on movement" to "constraint on slash transfer" may not appear to constitute a major difference, the SSH justifies the difference in terms of ease of processing and learnability. SSH argues that the original justification on constraints on wh-movement was derived from the assumption (Chomsky 1981) that if anything could move, it would move anywhere ("Move α"). Since such a rule would be wildly overgeneralised, constraints had to be identified (either language-specific, or universal) to disallow specific instances of movement. Thus generative grammatical accounts would result in multiple "specific technical constraints … as part of the syntax per se" (ibid: 337). However, under the SSH, it is suggested that generalisation is relatively conservative, and based on well-supported inferences about structures predictable from the input (e.g Tomasello 2003). Culicover and Jackendoff suggest a reversal of the "Move α" rule, that requires constraining, to a learnability assumption that starts with "no extraction from anywhere … and you then learn the possibilities in your language from positive evidence" (Culicover and Jackendoff 2005: 332). They tentatively hypothesise that the mechanism guiding learnability could be guided by a kind of hierarchy of processing complexity (along similar lines proposed by Keenan and Comrie (1977) for an Accessibility Hierarchy in relative clauses). Thus simple extraction would be easier to process than complex extraction, because of the difference in chain complexity. However, the complexity introduced by constraints is not part of the syntax per se but is rather stated in terms of syntax–CS correspondences. Specifically, chains contribute to "complexity in the mapping" at the interface between syntax and semantics (CS), because the head of the chain is displaced to the position marked by the operator (ibid: 334). Therefore more complex (longer) wh-expressions involve greater semantic complexity as well as greater syntactic complexity and take longer to process. However, according to SSH, the mapping between CS, SS and PS is all seen as contained with a language module, and therefore the role of memory, as part of the language module or as an extralinguistic cognitive resources, is not entirely clear.

Turning now to Head-driven Phrase Structure Grammar (HPSG), this theory is based more strictly within a processing-only approach, rejecting the concept of traces altogether. HPSG theory, like SSH, is a principled theoretical account of principles and structures, but sees syntactic structures as driven by surface processing constraints. This account can be seen in conjunction with cognitive psycholinguistic accounts of processing restrictions on island constraints. These accounts drive active research in L1 sentence processing, in which working memory plays a key role (e.g. Kluender and Kutas 1993; Gibson 1998; Phillips 2006). In this account, a fronted wh-phrase, or filler, is thought to trigger a process of searching for a matching gap (Frazier and Clifton 1989). The filler must be kept in working memory until that gap is found, which incurs a processing cost the further the gap is from the filler (Gibson 1998).

HPSG accounts of wh-movement (Pollard and Sag 1994, Ginzberg and Sag 2000) have focused on accounting for constraints on long-distance movement rather than the matrix question phenomena alluded to earlier (wh-fronting, subject-object asymmetry, no inversion in embedded clauses). In HPSG and other processing-based research, long-distance constraints on complex island extraction are governed by processing and memory limitations (Gibson 1998) rather than by responding to innate syntactic constraints. Hofmeister and Sag (2009), for example, provide evidence from experimental reaction time and neurological data to argue that some syntactic islands may be the result of aggregated processing difficulty including semantic encoding. They identify a difference between slower times taken overall on complex wh-phrases and processing times of the filler-gap retrieval sites. Contrary to intuition perhaps, they find that complex wh-phrases facilitate processing compared to single wh-words; that additional semantic and syntactic features increase encoding times, but facilitate processing at retrieval sites; that reading time benefits for complex NPs begin immediately after subcategorizing the verb; and that ease of processing matches up with higher acceptability judgments. I have not yet been able to find any evidence of research using HPSG theories in SLA, and the details of the theory are beyond the scope of this research study. Nevertheless the theory is mentioned here to clarify that processing constraints could well play a major role in a full account of the acquisition of L2 wh-movement, especially to explain asymmetries in judgements on different types of long-distance extraction. In addition, working memory is specifically assumed to constrain processing.

A third example of a formal but non-Chomskyan framework is Lexical Functional Grammar, or LFG (Bresnan 1981, 2002), a detailed explanation of which is beyond the scope of this study, but requires some mention, since it underpins a widely-applied processing-based theory of how questions may be acquired in SLA: Pienemann's Processability Theory (1998), which will be referred to later in section 7.1. LFG comprises a tripartite structure comprising a constituent (c-structure) component that generates constituents and constituent relationships at surface level, a lexicon with syntactically specified entries, and a functional (f-structure) component which acts as a processor compiling all the grammatical information (or features) needed to produce or interpret a sentence. The theory is constructed to avoid transformational operations such as movement, since lexical entries can be specified for multiple functions. For example, in a question such as "*What did you eat*?", *what* is specified both for question-focus and object, and can be parsed at sentence-level without recourse to underlying positions or traces (Bresnan 2002). Since the predictions of Pienemann's Processability Theory do not depend intrinsically on the theoretical constructs of LFG, I do not go into more detail here. Moreover, it does not appear that LFG is any more likely than other processing-based accounts to explain or predict potential asymmetries in acquisition than the other models described above, and is therefore not considered central to the issues investigated here.

2.3.2. Non-formal cognitive accounts

In comparison to the formal symbolic systems described above, there are other theories within a broad-ranging functional-cognitive framework which argue that language is part of general cognition: parsing the surface string or lexical input drives acquisition. There are a range of specific theories and models, but the main assumption is that linguistic knowledge of words and constructions is learned through inductive mechanisms, working on input through frequency, salience and ease of processing, and then stored in memory, via construction-specific rules (e.g. Bybee 1985; Goldberg 1994; J. Hawkins 1994, 2004; Elman et al 1996; Tomasello 2003; O'Grady 2005, 2008). In some of these theories, language, along with other memorised learning, is an amalgamation of subsymbolic representations or "connections", rather than relying on any kind of symbolic structures. Although memory is not usually specifically defined in terms of implicitness or explicitness, it appears that, in these accounts, much of language is explicitly learned as word or phrase-based constructions, and then become part of implicit long-term memory through practice (Tomasello 2003). Many of the phenomena that are assumed to reveal underlying specific syntactic principles, such as left or right-headedness are argued instead to be driven by the psychological parsing processes involved in language computation and use (Diessel 2004: 25).

An alternative to the pure "either/or" research paradigms of the generativists vs. the subsymbolic connectionists, is a growing interest in looking at language within a complex dynamic or "emergentist" framework, in which memory is central (O'Grady, 2005, 2008), which shares some similarities in approach with HPSG and SSH discussed above. O'Grady combines the assumption of the generativists (and others) that there are symbolic properties of language with the assumption of the cognitivists of the importance of processing in real time. His specific proposal is that "the core properties of sentences follow from the manner in which they are built" (2005: 2). He holds that syntactic phenomena are best understood in terms of a simple efficiency-driven linear processor resolving symbolic lexical properties. His understanding of symbolic properties is conceptually little or no different to the standard generative view of lexical categories and their combinatorial possibilities. However, he rejects what he sees as the generative idea of a distinct language module, in which the linguistic computational system also "incorporates a grammar" (ibid: 4). O'Grady proposes a dual system (see discussion in sections 2.4 and 2.8 below) distinguishing a conceptual-symbolic system for linguistic symbols and their meanings (both words and morphemes), which he associates with declarative memory. The computation system is a separate system, providing a set of operations for combining lexical items which "corresponds roughly to what we normally think of as syntax" (2005: 3) and is arguably an instance of procedural cognition. The computational system operates linearly, and is programmed to resolve any unresolved lexical requirements or dependences at the first opportunity (ibid: 5), but the information is held in working memory until it is resolved. In the instance of wh-questions (such as *What did John do?*), O'Grady argues that wh-expressions have an intrinsic dependency for meaning, e.g. between *what* and a verb without an overt object argument, which he terms an "unresolved wh-dependency," 2005: 113. Such dependencies are "held in working memory" (ibid: 115) and passing "downward" until a suitable dependency is located (ibid). The system is thus intended to "not only track frequency, create associations, and compute distributional contingencies" but also also "minimize the burden on working memory" in the course of sentence formation and interpretation (2008: 452).

To conclude this review of generative and non-generative theories of what linguistic knowledge is (or the property of language), all can agree on some aspects, but disagree on others. They all largely agree that, for native speakers, the knowledge base for question constructions, as investigated here, is largely implicit. They differ in whether this kind of knowledge is in a separate language faculty or not, and to what extent processing should play a role in how native speakers come to acquire this knowledge. There is a growing trend towards finding a path that combines the insights of crosslinguistic generalities found in generative linguistics with growing understanding of the role of processing mechanisms, particularly in terms of language use and acquisition, in an attempt to move away from the historical antithesis between generative and non-generative accounts of language. Sprouse et al (to appear), for example, draws together data from generative and processing-based studies to explain different effects in acceptability judgements between wh-extraction and wh-in situ. They conclude that both working memory constraints on online gap-filling and also feature-based licensing constraints can provide a fuller reconception of the operations of overt and covert movement. Hence they argue that a closer interaction between the syntactic literature and the sentence processing literature is not only possible but necessary.

This study is motivated by this cross-disciplinary trend in native speaker (L1) research to investigate research questions about how L2 language is acquired, particularly in terms of the different stages of acquisition (or transition), in terms that refer to both feature-based and processing-based accounts, as an attempt to contribute to this debate.

2.4. L1 acquisition

I turn now to a very brief review of theories of acquisition, in order to show how the various frameworks discussed above are argued to explain child acquisition from research perspectives which have been influential in SLA. The scope of this study precludes greater depth of analysis, but where appropriate, the theories and models will be revisited in section 6 on SLA.

2.4.1. Generative approaches to language acquisition

According to generative theory, the language module referred to in section 2.1 earlier, which contains the abstract principles that shape how features operate in any given language, is separate from but interfacing with the general human cognitive system. Following Fodor's (1983) characterising of modular processing, linguistic knowledge (or competence) is implicit or unconscious, and the computational procedures or processing mechanisms that interact with this unconscious competence are automatic and fast.

However, it is not always clearly explained how such competence and processing mechanisms are acquired by children. Two major assumptions underpinning generative principles in acquisition (see, e.g. Chomksy 1980; Hornstein and Lightfoot 1981; Clahsen 1996) are, firstly, Plato's argument of "poverty of the stimulus" (that what children hear is much less rich than what children can eventually creatively produce), and secondly, the logical problem of "no negative evidence" - that children cannot learn what is ungrammatical from the ambient grammatical input. Arguments have been made to show why input in itself is insufficient to guarantee acquisition, through the presence of non-relevant extra information or "noisiness" in the input, and the non-effectiveness of adult correction (Marcus 1993; Guasti 2004). A third assumption held by some (though not all) is that children and adults share greater similarity in their language systems than their output suggests (the Continuity Hypothesis, e.g. Pinker 1994).

However, if the grammatical system is the same for both children and adults, it has long been asked how language develops in children from early non-inflected output to full grammatical output (e.g. Pinker 1984; Hyams 1986). Generative researchers themselves do not agree (Clahsen 1996), but tend to follow one of several arguments. Firstly some hold that full competence is present from the start (the Continuity hypothesis) but that early omission of morphemes is due to external developmental constraints (e.g. pragmatic awareness, memory/processing limitations, learning mechanisms). Wexler's Maturational Hypothesis (1994) suggested that Tense features only mature around 2;6 years. Rizzi's Truncation Hypothesis (1993/4, 2000) takes this idea forward suggesting that some functional categories are initially present but truncated at TP until around 2;5 years due to immaturity of the production system.

Others adopt a structure-building approach (e.g. Vainikka 1993/4; Radford 1996), which has parallels in SLA theory (Vainikka and Young-Scholten 1994; Hawkins 2001). Radford (1996) argues that not all functional categories are available from the start, but children "build up syntactic structures one projection at a time" (1996: 43). He suggests that TP and CP structures develop later for a number of reasons. The first is that acquisition is lexically driven (ibid: 73). Words are learned before functional categories, possibly because of the lack of acoustic salience – much of the morphosyntax of tense inflection is argued to be non-salient, e.g. "What*'d* you do? What*'s* he done? He ask*ed* me the way". He argues that, as in Chinese, wh-expressions can be typed as either interrogatives or quantifiers, and that optionality ceases once children have learned on an item-by-item basis how to identify a wh-expression correctly as an interrogative.

On any generative account, it is assumed that development through the stages towards adult-like native speaker competence is "triggered" by abstract triggering mechanisms and learnability algorithms such as the "Subset principle" (Pinker 1984; Goodluck 1991; Gibson and Wexler 1994; Bertolo 2001; Fodor 1999; Lightfoot 1999; Fodor and Sakas 2005). Language acquisition is explained in terms of children making conservative hypotheses about what new grammatical rules might be needed to yield a possible interpretation. Learnability is argued to be a set of UG principles acting on the input to provide the child or learner "with an orderly procedure for positing hypotheses about the structure of the language" (Goodluck 1991: 150). The Subset Principle, for example (Berwick 1985), suggests that language development is driven by a learning mechanism that assumes the most restrictive grammar, and only changes functional instantiations in the grammar to the "smallest superset" of that language when faced with positive evidence in the input (Fodor and Sakas 2005). In other words, functional categories allowing increasingly complex structures to be produced (such as Tense) are acquired with only the minimal addition of features required to interpret input in line with innate grammatical principles.

Lightfoot (1999) discusses how the Triggering Learning Algorithm (Gibson and Wexler 1994) might work within the context of principles and parameters of underlying innate grammatical constraints on word order, headedness and so on. Grammar-language pairs of specifier and complement settings are argued to produce a finite set of possible word order settings (e.g. VOS, OVS, SVO, SOV), and input should conform to one of these settings. A child will react if their internal grammar cannot parse the input by adopting a grammar that "most resembles the grammar which generates that input" (Lightfoot 1999: 147). However, Lightfoot (1999) suggests that this is insufficient in understanding a wide range of evidence (e.g. development of creoles) and argues for a more "cue-based" approach to acquisition, in which increasingly complex and abstract cues are "some kind of structure, an element of grammar, which is derived from the input" (ibid: 149). In this sense his use of cues is tied strongly to the grammar, rather than the kind of semantic cue-based approach of some cognitive models (see section 4.3 below). Lightfoot argues that cues both are lexically based and input-driven, but also conform to innate knowledge of abstract phrasal categories and headedness. Therefore, children's early learning of certain analytic vocabulary cues the development of right or left-headedness; the later capacity to assign phrasal categories, functions and thematic roles is used to look for appropriate cues, e.g. to determine whether the specifier of CP is filled or not in developing V2 word order (ibid 174).

However, issues have been raised with generative models of child acquisition, such as those outlined above, that the nature of the developmental mechanisms or triggers suggested above are not always worked out clearly (Atkinson 1996; Tomasello 2000). In particular, how precisely non-linguistic factors such as maturation or increasing processing capacity interfaces with, for example, development of functional categories are not always specified, despite a long record of the need to consider processing and other constraints on child acquisition (e.g. Pinker 1984; Frazier and de Villiers 1990).

Some specific mechanisms that have been specifically posited to drive child acquisition include lexical learning and "bootstrapping" mechanisms (Pinker 1984; Bloom 1994; Guasti 2003). Children can segment the sound stream into distinct words from birth, and into clauses from around 6 months, using different sources of language-specific information as phonological "bootstrapping", such as prosody, distributional regularity and phonetic cues (Guasti 2003: 96). Children "innately know that words refer and that others use them with a referential purpose" (Guasti 2003: 96); there is argued to be a preference for noun learning, which may be due to word-world mapping procedures where the concreteness of nouns helps earlier learning of nouns over verbs (ibid: 97). Phonological bootstrapping is also argued to interact with semantic bootstrapping, where awareness of how words are used (e.g. nouns, verbs) drives the development of how to instantiate pre-existing UG principles in the child's specific language (Pinker 1984). The role of online processing or "operating principles" have also been argued to play a key role in how children learn to disentangle form and meaning from the input (Slobin 1985; Clark 2003).

Discourse-level issues, such as interaction and awareness of pragmatic meaning are also seen to be crucial to child development. Interaction is argued to be required (Snow et al 1976; Clark 2003), in that ambient language addressed to others has not been found to trigger development in the way that child-addressed language does; ensuring joint attention and scaffolding comprehension (grounding) are seen as necessary elements of the input. Van Valin (1998) argues that pragmatic awareness develops early in children, and that this is sufficient to drive their awareness of how questions can be formed (1998: 235). He highlights subject-object asymmetry in the order of acquisition of questions, building on Stromswold's (1995) findings. In terms of syntactic complexity, subject questions should be acquired first, since there is no need for verb-raising (subject-auxiliary inversion, or do-support) but Stromswold found no evidence for greater comprehension and production of subject questions (1995: 16). Van Valin

(1998) argues that this is best understood within a framework of looking at pragmatic focus. He claims that all languages have a universal "unmarked" or default predicate focus structure (ibid: 237), and in English, this is found in the post-verb position. Hence object questions are thus easier to process in English, since they correlate with the unmarked post-verb position. A similar pattern was found in long-distance extraction where object questions emerge first, again reflecting pragmatic focus differences, as well as syntactic considerations (since object questions are not constrained by the *that* trace effect) (Van Valin 1998: 237). Van Valin also discusses the evidence of intermediate copy wh-expressions referred to above (in section 1.2.3) in pragmatic terms of children attempting to distinguish where the actual focus lies (ibid: 238). He supports Thornton's (1995) observations that these copy or medial wh-expressions disappear from object and adjunct questions during development, since these are easier than subject questions to assign actual focus (ibid: 240).

There is thus a continuum between the strongest "pure UG" account of child acquisition driven by abstract algorithms, to other accounts which may be based on UG syntactic features, but which also incorporate elements of processing, bootstrapping and pragmatic mechanisms. However, there are other "pure" cognitive accounts which assume acquisition is driven wholly by processing constraints and general cognitive learning mechanisms.

2.4.2. Functional-cognitive accounts

Non-generative accounts of acquisition are specifically based on processing and learning mechanisms, thus intrinsically incorporating issues of learnability within general cognition.

Functional-cognitive approaches, as outlined in section 2.4.2 above, see linguistic knowledge and learning mechanisms as located wholly within the general cognition system and driven by the input (Anderson 1983; Bates and MacWhinney 1989; Goldberg 1994; Bialystok 1994, 2002; Tomasello 2000, 2003; Diessel 2004; J. Hawkins 2004). Grammars develop through learning mechanisms which map form to function and meaning (Bates and MacWhinney 1989: 9), and child grammars are, in contrast to generative accounts, highly constrained in both representation and production by what children have learned. Children are argued to show conservatism in their use of novel forms until the age of around 3;6 (Tomasello 2000), by which point they have amassed sufficient item-based information, based on imitation of input, to drive more efficient analogy-making and structure combination mechanisms (ibid). Learning is driven by regularity and frequency of a structure in the input, and the ability to respond to phonological, semantic and morphological cues in the surface form, based on, for example sound, stress, role (e.g. agency), form, agreement, case, word order allow the language learner to infer underlying meaning or intention (Bates and MacWhinney 1989: 37, 42).

More recent accounts, such as the "constrained statistical learning framework" (Saffran 2003) respond to the evidence in generative accounts of crosslinguistic universals, by positing universal statistical and probability inferencing strategies for learning, highlighting the ability of even very young infants to track "sequential probabilities" from the statistical properties of the input language (ibid: 111). Knowledge of probabilistic word and sound sequences is stored in pre-packaged chunks or schemas, at phrase level (Tomasello 2000) or even narrative level (Rumelhart 1975). In sum, no information is processed and encoded that is not directly derived from the input by general cognitive mechanisms (Williams 2005).

However, cognitive accounts are criticised on conceptual and empirical grounds for lacking the universality and predictability of generative accounts (e.g. Pinker and Prince 1988; Marcus 2001).

The emergentist framework (O'Grady 2005, 2008) argues that language acquisition can be explained in a way that produces the universality and predictability of generative accounts without relying on an abstract Universal Grammar that must mature, or on parameters to be set. Instead O'Grady proposes that language acquisition or learning is "the emergence of computational routines" (2005: 193), which work to ensure maximum efficiency in interpreting the input with minimum processing cost (although issues of avoiding costliness in processing are also noted in more standard generative-based work from at least the 1980s, see e.g. Pinker 1984). O'Grady illustrates, as an example of this, the emergence of understanding how to interpret agent assignment on verbs, which he argues emerges from repeated instances of word-based learning of dependencies from highly frequent, very familiar items (2005: 195). The computational principle, for example, that transitive verbs interpret or find their dependent argument by "looking right" are learned through repetition of simple verbs such as *hug, kiss, eat, drink*. Later computations such as auxiliaries and copula verbs looking right for their argument in yes-no questions are learned later.

The emergentist account is very new, and has not yet widely been applied to other child acquisition research other than English, as far as I am aware, and is not yet widely adopted within SLA research. However, in the prominence given in this account to the role of WM in language acquisition (both L1A and L2A), it is of particular note for the assumptions underpinning this research study. A recent special edition of *Lingua* (2009) set up a debate between the predictions of emergentist research compared to generative research in SLA, which is discussed in more detail at the end of section 6.11.

To conclude this discussion of child acquisition, it seems that there are problems and difficulties with too extreme a view of either generative or cognitive approaches to acquisition. Generative approaches are argued to need greater incorporation of non-linguistic learning mechanisms; cognitive approaches are criticised for lacking a broad scope of predictable crosslinguistic universality. In reality, as stressed by Clark (2003) and others, there are a wide range of factors that combine to assure successful L1 acquisition, most if not all of which can legitimately be equally applied to SLA.

Herschensohn (2007: 63) writing about both L1 and L2 acquisition, states:

> "Native language acquisition…requires input from the ambient language; draws on innate predispositions at every stage; exploits linguistic, pragmatic, social and environmental scaffolds; uses prosodic, semantic, syntactic and lexical bootstrapping; calculates frequency and saliency of input; and completes the process by creating native competence in grammar".

Herschensohn's identification of at least twelve factors affecting acquisition for children makes it very clear just what a complex task L2 acquisition is, even if some of the maturational constraints no longer apply to adult learners.

It is also clear, from the discussion above of theories of both linguistic representation and language acquisition, that language use, or processing in real time, should be taken into account to shed more light on acquisition of an L2. I turn now to present an overview of some models of mind that have been widely adopted in native language processing accounts, which have been influential in SLA research, in order to provide the final dimension driving the research questions of this study.

2.5. Models of mind

The two broad approaches to acquisition outlined above are represented by a distinction between two conceptualisations of models of mind based either on a separate language module interacting with other cognitive processes, or as language based within general cognition.

One of the most widely used models of mind adapting Fodor's conceptualisation of modularity is that of Smith and Tsimpli (1995). Their influential study of Christopher, a multilingual savant whose capacity to acquire up to twenty different languages was unimpaired despite general learning difficulties, was seen as an important source of evidence for an independent language module.

In their model, the language module is separated from the central cognitive system where knowledge is stored. There is some interaction with the central non-linguistic system, via the lexicon at the morphology and conceptual interface, allowing the language module's combinatorial rules to operate on stored knowledge of form and lexis to create grammatical language. Language processing is handled by parsing mechanisms which are assumed, in this approach, to be part of the innate language module, and therefore not part of the processing systems that interact with memory (Fodor 1999).

However, how the interface or computational processing mechanisms work is not always clear. In more recent years, generativist accounts of language acquisition (Hauser et al 2002; Chomsky 2005) have simplified and extended the way in which stored knowledge and combinatorial principles operate, opening the way for a closer integration between implicit linguistic competence and general cognition. While the abstract principles operating on lexical information remain modular and separated from other implicit processes within the mind, morphological and syntactic features are now seen as attached within the lexicon, which overlaps with the general cognitive system more in this definition than the Smith and Tsimpli model.

2.5.1. Parallel Architecture

Jackendoff (1997, 2002, 2007), building on the above change, suggests a model that more closely links lexicon, memory and processing. Jackendoff's Parallel Architecture, along with his Simpler Syntax Hypothesis, aims to provide an answer to the question of how a linguistic theory of grammar and structures can be linked to a theory of processing language in real time – the "logical problem of language processing" (Phillips and Lau: 2004: 16). In this model (Figure 1: Jackendoff's Parallel Architecture, below, adapted from Jackendoff 2007: 8), there are three independent generative components for phonology, syntax and semantics, linked by interfaces.

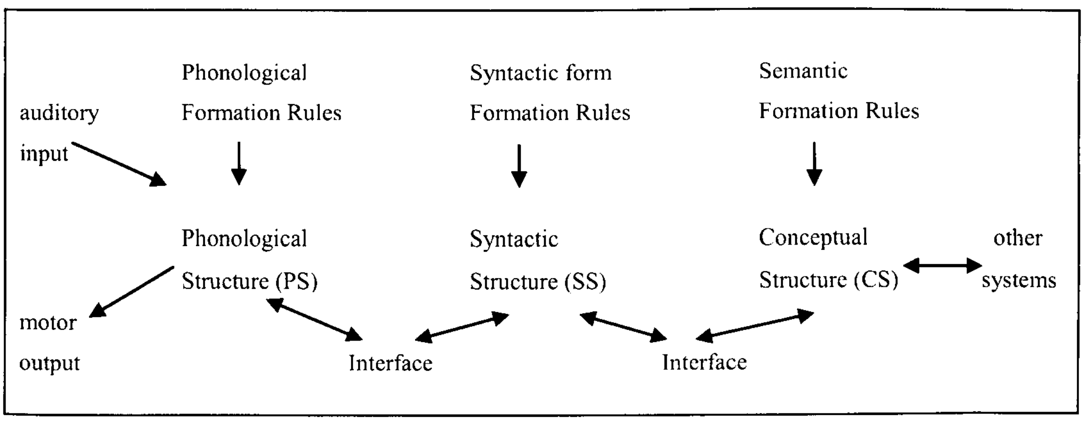
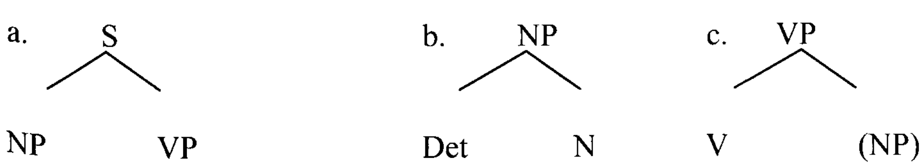


Figure 1: Jackendoff's Parallel Architecture (2007)

Auditory input and motor output works initially on the phonology component for perception (for production, the direction of input and output is simply reversed). Each component has its own set of combinatorial or formation rules. Grammatical structures can be composed from pre-formed prototypical "treelets" (as shown below).

Prototypical treelets:



A tree can be built by "clipping together" these treelets (ibid: 8), working bottom up or top down. The structures thus act as constraints on possible trees, rather than as "algorithmic generative engines" for producing trees bottom-up. The knowledge of possible, even probable, tree structures aids incremental parsing. Jackendoff (2007) points out that this kind of constraint-based formalism is shared with other formal non-Chomskyan approaches to grammar such as Head-driven Phrase Structure Grammar (mentioned above), but differs from purely lexically-driven connectionist architectures (see above), in assuming that structure-building is constrained by a "mixture of word-based, phrase-based, semantically and even pragmatically-based conditions" that act in parallel (ibid: 9). There is no strict word-rule distinction, since both contentful structures, such as words and formulaic phrases, and contentless structures such as preformed treelets outlined above, are "simply different sorts of stored structure" stored in long term memory in the lexicon (ibid: 11).

In order to parse auditory input, Jackendoff posits a separate linguistic working memory as a temporary workspace or "blackboard", differentiated for phonology, syntax and semantics (ibid: 14), in which the incoming information can draw in parallel on information stored in the lexicon to build an incremental parse.

Jackendoff thus retains UG-constrained symbolic systems, located within memory, but with language-specific processors which interface between learned knowledge of words (lexicon) and input and output, and with language-specific working memories. The model is, I argue, very helpful in moving on from an overly abstract constrained notion of modularity in which processing or general cognition plays no role, and specifically addresses the question of how to incorporate evidence from processing studies. The model is also parsimonious in positing the same system for processing input and output. However, it is open to questions in its definition of working memory and the lexicon, particularly as to how the mechanisms of explicit or implicit learning or storage may be differentiated. He does not discuss a procedural-declarative distinction in long term memory, which makes his model look similar to older conceptions of memory (Atkinson and Shiffrin 1968). He also does not raise the question of awareness, automaticity or control. A third problem is that Jackendoff has not yet spelt out how this architecture explains L1 acquisition. Despite these shortcomings, the model has been influential for two significant models for L2 acquisition – Carroll's Autonomous Induction Theory, and Truscott and Sharwood Smith's MOGUL model, which I will refer to later in section 2.8.2.2.

2.5.2. Competition and connectionist models

Alternative accounts place language processing wholly within general cognition, such as Bates and MacWhinney's Competition Model (1989), connectionism (McClelland and Rumelhart 1986; Elman et al 1996) and emergentism (O'Grady 2005, 2008). These accounts have been influential within cognitive approaches to SLA, particularly the Competition Model, which is closely associated with the emergentist framework (e.g. MacWhinney 2001, 2004; O'Grady 2008) in a drive to present the general framework as a unified model of L1 and L2 acquisition (MacWhinney 2004).

The influential Competition Model suggests that grammars are sets of rules created in response to the strength or "weight" of cues in the input, which provide a set of "partial solutions" to the mapping problem between form, function and meaning (Bates and MacWhinney 1989: 9). Different possible parses compete between these partial solutions to yield optimal comprehension. The parsing process itself is also subject to competition from the "different and competing demands of perception, production, learning and memory" (ibid: 25). Language learning is thus "a process of acquiring coalitions of form-function mappings, and adjusting the weight of each mapping until it provides an optimal fit to the processing environment" (ibid: 59). The memory system itself lies outside the cognitive content of grammatical forms (ibid: 24) and is separate from the learning system, and both put pressure on the cognitive system as a whole in terms of processing ease or resource availability (ibid: 25). In this approach, registering or noticing of the input may be implicit or explicit, but conscious registering of cues in the input may aid the development of explicit linguistic knowledge. It is logical then that this model has been taken as central in analyses in SLA of the role of instructed learning, and the role of awareness in SLA (Schmidt 1990; Robinson 2001).

Connectionism provides a context for drawing on the ideas behind the Competition Model for a specific model for learning. It derives from models of the way the brain is argued to build up knowledge through millions of simultaneous computations, in which densely interconnected networks of input and output units act in parallel and build up "knowledge" through gradually increased strength of connections between units (McClelland 1989: 427). Such networks, it is argued, can be trained to produce human-like responses on word recognition and simple syntactic rules like English past tense, suggesting that they are able to extract knowledge of underlying regularities, rather than merely storing the specific items received during training. Evidence has also been found that such networks can predict word boundaries in speech and word classes (ibid: 3); it is also argued that they can make structure-dependent predictions over long distances, and show sensitivity to embedding and recursion, suggesting that they could form the basis of a connectionist approach to syntactic processing (ibid: 4). Rule-like behaviour thus emerges, not as a result of underlying symbolic forms, but subsymbolically, or atomically, as a consequence of "the way in which individual instances of experience are stored in a single memory system" (Williams 2005: 4).

However, issue has been taken with pure connectionist models (e.g. Pinker and Prince 1988; Marcus 2001; Pinker 2004; Jackendoff 2002, 2007). Criticisms are raised over evidence derived from "small vocabularies and a small repertoire of structures" (Jackendoff 2007: 13), and the conceptual difficulty in responding to evidence of underdeterminism in the input and lack of negative evidence (Pinker 2004). It is unclear that connectionist models could manage or predict the complexity of natural language, by the full interpretation of a novel syntactic, semantic and intentional utterance such as "I'll meet you for lunch at noon" (Jackendoff 2007: 13) or "general relations" such as semantic identity, rhyme or matching irregular past forms to present forms (ibid). In addition, one of the principal difficulties with connectionist models, as with other cognitive accounts discussed above, is that they do not specify precisely how linguistic representations are stored as networks in memory, and particularly, how implicit knowledge, which is the basis of the networks referred to here, interacts with information about language that could have been explicitly taught, such as word meanings or more complex forms such as passives and relative clauses.

Cognitive accounts are also not always clear in distinguishing processing constraints from memory constraints, especially in terms of how far parsing involves any degree of conscious awareness, and whether working memory is involved in dealing with the "processing load". There has been much discussion of the different accounts of extraction constraints on long-distance wh-constructions, but I have also not been able to establish how far cognitive accounts may also provide equivalent arguments for the crosslinguistic differences on Tense, head movement, do-support and other phenonema discussed from a formal perspective earlier.

Cognitive and generative accounts, as mentioned at the start of this section, seem to differ from each other in their understanding of how implicit knowledge develops, and whether there is a role for conscious awareness in initial learning of words and phrases. This relates to an assumption in many theories (though less obviously in "pure" generative work) that there is a distinction between two types of linguistic knowledge – a Dual Mechanism typified in Pinker's "Words-Rules" paradigm (1999) distinguishing rule-based automatic constructions of regular past tense formation, from retrieval of irregular past tense forms. This distinction is potentially crucial for children who are using learned lexical knowledge to drive syntactic development through semantic bootstrapping. As seen above, both generative and non-generative accounts assume that mature linguistic knowledge leading to use of regular "rules" is non-conscious, or implicit. Knowledge of the meaning of words in the lexicon, such as the form of English irregular past tense forms, on the other hand, is argued to represent a different kind of knowledge, which stems from explicit learned or memorised items. However, Pinker makes the point that in some word-rule situations, such as in the case of past-tense generation in English, inferring a strong dichotomy between implicit rule generation and explicit word retrieval is too simplistic. Pinker argues that while "irregular inflection is inherently linked to memorized words or forms" (1999:10), there is an additional element of implicit computation required for production of irregular verbs (ibid). Nonetheless, there is some extent to which it could be argued that memory works in different ways for regular rule generation than for production of items inherently based on explicit learning.

In this discussion of L1 acquisition, I have argued that in order to understand how language is represented, acquired and used, both formal and processing accounts can be considered. It has also been shown that linguistic knowledge and memory are seen in different ways in different frameworks, particularly in the apparent dichotomy between implicit or inductive knowledge of rules and explicit knowledge of words. It is therefore crucial to consider in more detail how memory is seen to operate in language, so that this apparent dichotomy between explicit and implicit linguistic knowledge can be better understood, both in terms of how linguistic knowledge is stored, as well as mechanisms of using that knowledge in real time.

I turn now to provide a brief summary of the standard psychological constructs of memory, storage and retrieval. I show how these are understood to operate for language development and use to differing degrees of implicitness, automaticity or conscious control, focusing on models and hypotheses that have been particularly important in theories of SLA.

2.6. Memory, storage and use

In standard psychological presentations of memory (e.g. Squire 1992; Smith and Kosslyn 2007; Baddeley et al 2009), there are generally accepted to be two types of memory – long term and short term or working memory. Within long term memory, there are held to be two types of knowledge store, divided into "knowledge about things" (declarative) and "knowledge how to do things" (procedural). Declarative memory, also called explicit memory, is seen as subserving information that can be represented consciously. The declarative memory store is generally agreed to consist of two types of knowledge: semantic knowledge (or encyclopaedic factual knowledge of the external world), and episodic knowledge of specific events in time (Tulving 1983), also known as autobiographical or experiential memory (Penfield and Roberts 1959). Declarative memory is highly flexible and items can potentially be stored very quickly (hearing a name once may be sufficient to store it), and is required for most kinds of lexical learning, even by very young children. However, accessing declarative memory is deemed to require costly serial processing and thus can result in slower overall processing as more items or more complex combinations of items are required, such as recalling how to create a grammatical construction by applying learned rules (Paradis 2004).

Non-declarative or implicit memory is less easily defined, and there is some debate in the psychology literature over its precise nature (Ullman 2005). It is generally presented (Anderson 1983; Cohen et al 1985; Fabbro 2002; Gupta and Cohen 2002; Paradis 2004; Smith and Kosslyn 2007) as comprising different sub-types of implicit knowledge: procedural, priming and conditioning. These different subtypes are accessed using unconscious processes or procedures. Procedural memory (for habits and motor skills) arises from repeating an activity, such as tying shoelaces, finding our way around a familiar room in the dark, riding a bicycle, learning to ski, playing a musical instrument. Some of the unconscious information which drives the activity at early stages may also be stored in parallel in explicit form (as in a beginner skier, who may ski slowly and effortfully down a hill, consciously repeating the teacher's instructions how to turn, lean, stop and so on, but who could, after a few lessons, descend the slope using unconscious procedural knowledge of the movements required, while thinking consciously of something else). Another type of implicit knowledge, or priming, is the process whereby presenting an item influences the processing of a subsequent item, even when no information about the subsequent task is given or drawn attention to. An example of this could be when we are shown or hear a set of words, and then are asked to name as many random words as possible, we are more likely to generate items already processed (Baddeley et al 2009: 82). A third type, conditioning (associative learning), is learning from experience, e.g. when a tone is played before a short puff of air to the eye, we "learn" to blink in advance just by hearing the tone (ibid: 12).

It is believed that implicit memory precedes explicit memory in children, perhaps because the neural basis for implicit memory is more developed at birth (Baddeley et al 2009: 290); also, implicit memory has been found in experiments to be superior in children of 3 years of age in comparison to their explicit memory (Paradis 2004: 9). Implicit memory does not typically show age-related changes in children, in comparison to explicit memory, which develops significantly over time from infancy and throughout the teenage years, perhaps facilitated by improvements in working memory capacity from the age of around 4 years old, as well as memory strategies and actual knowledge based on experience (ibid: 274). However, even by one year old, infants have been found to show evidence of explicit memory, learning information quickly, being able to remember and apply learned information, suggesting that the traditional assumption that infants had little capacity for explicit knowledge needs revising. During normal adult life, declarative episodic memory of events remains constant from young adulthood (around 20 years old) to around 50 or 60 years old; declarative semantic memory for words is maintained, although speed of access declines around a similar age (ibid: 302). Implicit memory and learning show some age effects, e.g. in testing how successfully people can complete word fragments, where letters are missing (ibid: 303), motor performance declines with age but it is arguable that the rate of motor learning is not necessarily age-affected, e.g. in training how to use a computer mouse to navigate a maze on-screen can be very successful even in over-50s (ibid).

2.6.1. Working Memory

The construct of working memory is also seen to be an essential part of the cognitive system, underlying "the maintenance of task-relevant information during the performance of a cognitive task" (Miyake and Shah 1999: 1). The separation of some kind of immediate or short-term memory as distinct from long term memory has long been articulated (such as the notion of primary attention, James 1890) to account for items we can hold in conscious mind for a short time, using controlled attention or awareness. If we are distracted, our attentional control cannot hold on to the information. Early accounts focused on the limited capacity of a short-term storage, following the work of Miller (1956), who identified a specific limitation in storage lasting around 1-2 seconds, which can contain a finite number of items, argued by Miller to be around seven (plus or minus 2). Telephone numbers, postcodes, and other items that require memorisation, are obvious examples of items that obey this limitation. Later research identified that this finite limitation was not confined to a strict countable number, but that if information could be linked together through a common functional context (for example into strings, sentences, or other connected "chunks"), then the limit of seven, applied to such chunks, could allow for much longer memorisation.

Further empirical evidence for a separate construct of short-term memory came from work with amnesics such as Scoville and Milner (1957), culminating in Atkinson and Schiffrin's model of memory (1968), suggesting that short-term memory was the "gateway by which information can gain access to long term memory" (Smith and Kosslyn 2007: 247).

This led to a widely held view that short-term memory was separate from but connected to long term memory (LTM), and that long term storage was dependent on conscious attention which acted as a filter (Atkinson and Shiffrin 1968), limiting the amount of information entering or remaining in the temporary store. However, it was not always clear how far short-term memory was defined by the notions of attention and consciousness, or was a separate construct in some way, as Atkinson and Shiffrin's model suggested (Jonides et al 2008).

It also became clear, in further research, that the assumption that whatever was consciously attended to would be stored did not always hold. Items could be repeated consciously many times and not recorded in long term memory; other information could be processed briefly but successfully stored – in other words, what mattered for long term storage was not amount of conscious processing but the level at which information was processed. Craik and Lockhart (1972) evaluated the difference between shallow processing (form, colour, shape, sound) and deep processing (meaning, implication), concluding that short-term memory was connected to long term storage but the means of processing was more important than short-term capacity itself.

However, the early construct of working memory as "gateway" or bottleneck on learning was influential in early work investigating its effect on L2 learning and processing (Lado 1965, Kolers 1966, see also Brown and Hulme 1992).

2.6.2. Baddeley's multicomponent model

A conceptual change from seeing short term memory as a single "gateway" to long-term memory (Smith and Kosslyn 2007) was suggested in the 1970s, after research on amnesics found that there could be different types of short-term memory impairment, which argued that phonological and visual storage, and attentional processing resources could be separated into different but overlapping components (Baddeley and Hitch 1974). Rather than simply referring to a single short-term store, Baddeley's (1986) multicomponent model of working memory (WM) identified the separate role of a "supervisory attention system", or central executive (following Norman and Shallice's (1986) model of attentional control) operating in conjunction with two separate storage components. As currently formulated (Baddeley 2000, 2003, 2007), this model posits that information coming into attention is temporarily stored in two domain-specific temporary storage buffers. Auditory stimuli are stored in the "phonological loop"; visual stimuli are stored on the "visuo-spatial sketch-pad". These stimuli are controlled by the central executive, which links to long term memory via the episodic buffer – see Figure 2 below.

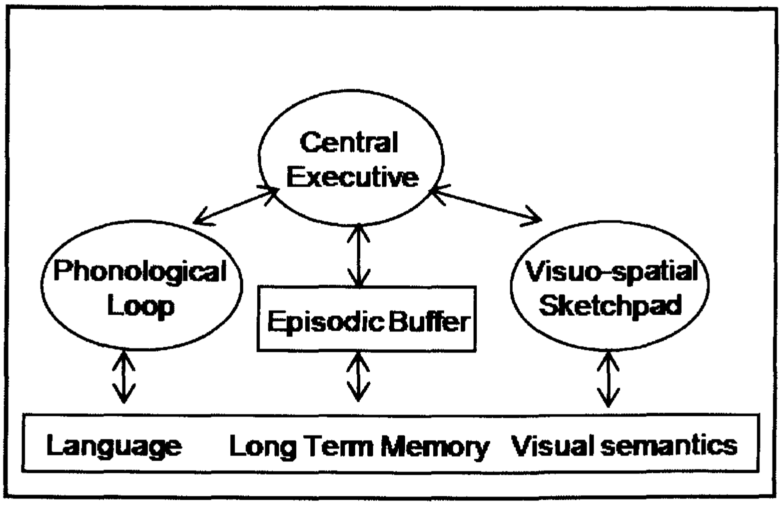


Figure 2: Baddeley's (2000) Multicomponent Model of Working Memory

The phonological loop is analogous to short-term verbal storage of between 5 and 7 items, where we can temporarily maintain verbal material such as a telephone number (as a trace), in the gap between looking up the number and then dialling it. Without rehearsal, the trace decays in around two seconds. The process of rehearsing (repeating internally or under the breath, usually called "sub-vocal articulation") allows for longer retention, while attention is focused on that task. Suppressing that rehearsal, for example by repeating a sound such as "the", has been used as a technique to measure simple storage capacity before the trace decays, and is usually specifically referred to as phonological short-term memory (PSTM), without a processing component. The visuo-spatial sketchpad is another system for storing visual material, such as spatial patterns or when we "picture" an item or place, either by retrieving it from memory, or by hearing a novel description.

The information stored in these two "slave" buffers, which are domain-specific, is managed for maximum processing efficiency through a "central executive", which intersects with long term memory via the "episodic buffer". The central executive provides two key functions: control and inhibition. Control is required to maintain and process the information required to complete the task, and inhibition ensures that the focus on task is not distracted.

The episodic buffer is a new element (Baddeley 2000), designed for temporary access to information held in long-term memory; in its interface between long term storage and central processing, it facilitates recall for longer than the standard 1-2 seconds, such as when recalling prose passages, and allows multi-modal conceptualisation requiring both visual and semantic activation (e.g. describing an elephant playing ice-hockey).

In this model, WM serves as a "workspace" both to store incoming information and retrieve stored information in order to complete a complex task, and is usually understood to operate when information involved in completing the task is consciously controlled. Therefore consciousness and awareness are central elements of WM. There are some limitations on Baddeley's model, particularly in the lack of detail in how language storage and processing accounts from linguistic and psycholinguistic research would fit with this concept of WM and LTM. In particular, Baddeley's model does not specify in detail how WM links with different constructs of declarative (explicit) or procedural (implicit) memory, which then makes it unclear how WM in theoretical terms operates for different aspects of language which are argued to be either automatic (and implicit), or controlled (and explicit). It is also not clear how this model can account for differences in bringing existing knowledge from long-term knowledge into the workspace, and how implicit knowledge may be utilised differently from explicit knowledge in being retrieved into working memory, or involved in working memory task completion, whether for language or indeed for other complex cognitive tasks.

Nevertheless, Baddeley's model has been the most influential in the language processing literature (see section 2.6.5 below), but other constructs have also been proposed, which address in more detail how short-term attention interfaces with long-term memory.

2.6.3. Other views of short-term and working memory

A second approach is that of Cowan (1999, 2005), who maintains a more unitary view of short-term memory in contrast to Baddeley's multicomponent model. Cowan's "embedded-processes" view of working memory is a subset of information from long term memory "held in a temporarily heightened state of activation" (1999: 63). Activated memory has a further subset of information which is the "focus of attention". This tripartite construct of degrees of activation and attention focused on items retrieved from long term memory is controlled by a central executive. Cowan also suggests the concept of "virtual working memory" which consists of those elements of long term memory associated with whatever is specifically activated, and which seem to be accessed faster. This suggestion overlaps with Ericsson and Kintsch's (1995) model of long-term working memory (LT-WM) which consists of developing specific areas of expertise in long-term memory which are encoded for swift efficient retrieval on demand as if they are in working memory. The emphasis on some element of central control is also discussed by Engle et al. (1999). Working memory capacity in this approach is taken to refer to "the capacity for controlled, sustained attention in the face of interference or distraction" (ibid: 104). This capacity is argued to equate to Baddeley's central executive, and is defined by Engle et al. as "general fluid intelligence" (1999: 126).

The distinction between Baddeley's model (separate from long term memory and multicomponent) and Cowan's model (currently active portion of long term memory and governed by attention) reveal that there is much disagreement amongst psychologists over how to define the construct of working memory. One of the major differences lies in the construct of the "central executive" which Baddeley admits is not yet sufficiently theoretically or empirically identified but is merely an unspecified "homunculus" (1999: 39) or a rather "insubstantial ghost" (2009: 148). A different conceptual approach takes a much more domain-specific view of working memory in relation to language, positing a specific verbal working memory construct, put forward by Caplan and Waters (1999). They follow Shah and Miyake (1996) in fractionating out the processing element of WM (the central executive) into verbal and visual components, and then into further sub-fractioned divisions between different types of verbal processing. Caplan and Waters suggest that the "interpretive processing" system (1999: 78) for assigning syntactic structure and meaning to a sentence is different to other verbal WM systems as used for long term storage, planning actions and other "post-interpretive" processing. This approach has been used in a few studies (e.g. for L2, Sunderman and Kroll 2009), but has not been central to much of the research in language processing and WM.

Miyake and Shah (1999) brought together ten theoretical but different viewpoints on the construct of WM. Their seminal debate challenged the leading researchers in the field to draw out the commonly agreed findings on WM. The resulting common agreements were that WM should not be considered as a separate "box" (Miyake and Shah 1999: 445) for short term storage that is structurally distinct from other memory systems. Thus the traditional view of a WM task of temporarily memorising an unfamiliar telephone number is unhelpful. Instead WM is defined as a set of processes in the service of complex cognitive activities, including language learning and processing. Therefore WM is less about memory and more about "control" of cognitive actions, as articulated by Baddeley in the role of the central executive and in other models by other mechanisms. WM remains limited in capacity, but multiple factors (including information-decay, processing speed, levels of knowledge or skills) can limit capacity, reducing the importance of attempting to define a single limiting factor (ibid: 448). There is an integral role of long-term knowledge and skills in WM performance, as articulated in Ericsson and Kintsch's (1995) LT-WM construct. Cowan's unitary model and Baddeley's update of the multicomponent model, by adding the episodic buffer interacting with LTM, are therefore closer to each other than in their original conception.

2.6.4. LTM and WM and language

In terms of native language, syntactic knowledge is argued to be largely unconscious; accounts of how native grammar is used, as outlined above, are thus typically situated in implicit or procedural knowledge and memory. However, it is difficult to find clear models of how memory is specifically involved in models of language acquisition, rather than language use (as outlined above). Generative accounts by definition assume that language acquisition operates within a separate language module, and that abstract linguistic representations are largely unaffected by memory constraints except in "performance". Cognitive theories of lexically-driven learning imply an interaction between explicit declarative and implicit procedural memory, but this is not always clearly drawn in the literature. Early psychological models of implicit learning based on artificial grammars (e.g. Reber 1967) have not been widely accepted as necessarily relevant to natural language acquisition (Brooks and Vokey 1991). Connectionist models refer to a "single memory system" but do not relate the mechanisms of developing connections to how language is acquired. Even avowedly usage-based accounts of L1 acquisition, such as Tomasello (2000, 2003), make little mention of specific memory mechanisms in L1.

If the role of LTM is not always clarified in the literature, WM constraints on processing have also been alluded to without always specifying how that would affect acquisition, since most research has been on adult native speakers (Harrington 2001). Some specific sentence processing research has assumed that working memory limitations affect successful processing of more complex structures, especially subjacency-constrained structures (e.g. Clifton et al 1994; Gibson 1998) and object/subject asymmetries in long-distance extraction, ambiguous embedded clauses or garden path sentences (Miyake et al 1994). Such findings have prompted research into such asymmetries in L2A, especially the object/subject asymmetry (e.g. Schachter and Yip 1990; Juffs and Harrington 1995; White and Juffs 1998), in the view that sentence processing research could add greatly to developing "a transition theory of SLA" (Harrington 2001: 124).

Yet, even in L1A, it is not clear how research into language acquisition and processing by linguists or psycholinguists interfaces with research into working memory driven by psychologists, leaving some questions about how language, LTM and WM all fit together. These cannot all be answered here, but I turn to research which has specifically investigated Baddeley's model of WM, in which WM has been found to have a robust impact on L1 language acquisition and use.

2.6.5. WM research into language acquisition and processing

Current WM research using tests for phonological loop storage and central executive efficiency has found robust evidence for the role of WM for certain elements of language acquisition and use in children.

In L1 studies of children beyond infancy (usually from four years of age upwards), measurements of WM have commonly been done via the phonological loop through word or non-word repetition (Gathercole and Baddeley 1993) and other measures of verbal, digit or visual recall and manipulation (now standardly produced as an automated WM assessment, Alloway 2007). Such tests have shown robust correlations with reading and vocabulary development in children, which require explicit learning. Phonological loop capacity has also been associated with oral production of longer, more syntactically complex utterances and the amount of story information recalled in a narrative retelling task (Adams and Gathercole 1996), which suggests that working memory plays some role in more complex morphosyntactic production, even though the assumption in much L1 acquisition research that grammatical acquisition is essentially implicit, and is largely in place by the age of four. However, to my knowledge, there has not been much research into WM in children and specific measures of acquisition of more complex forms of later acquired grammar, which are argued to require greater processing capacity (such as passives, or, as relevant here, long-distance questions). Research into WM in children of school age in developing reading and comprehension proficiency has had wide application, leading one leading WM researcher to claim that WM is "the new IQ" (Alloway 2009). Other studies have identified that language impairment or variability in early language development is closely allied to phonological and verbal memory difficulties (Speidel 1993, Adams and Gathercole 2000). Nevertheless, it is noted that most of this research has focused on learning of explicit knowledge (vocabulary, reading), and Gathercole and Baddeley (1993) do not assume that WM plays any great role in normal pre-school acquisition of morphosyntax.

Other research has focused on the role of WM in adult language use, drawing on Baddeley's view of a trade-off between phonological storage/central executive processing. This is where the two elements of the model (storage in the buffers, and processing by the central executive) are seen to trade off against each other – the greater the amount of information to be stored in one of the slave buffers, the less efficiently the information can be processed by the central executive (Daneman and Carpenter 1980). Strong correlations have been found between measures of this trade-off and reading proficiency (measured as the capacity to correctly identify or maintain discourse reference for pronouns which are distant from their referents), and lexical and syntactic disambiguation in sentence processing (Gathercole and Baddeley 1993; Miyake et al 1994; Baddeley 2003).

Daneman and Carpenter (1980) designed a reading span test for adults to test the limits of storage during linguistic processing in their native L1. University students were asked to read sets of sentences, grouped first in pairs, then threes, then fours and so on. After reading each group, the final word from each sentence was to be recalled. Capacity was measured in terms of accurate recall of the greatest number of words in correct order. Their research found strong correlations with correct pronoun reference assignment in a reading comprehension test. WM capacity as measured this way was also found to account for individual differences in resolving syntactic ambiguity in garden path sentences, when comparing potential subject or object assignment to relative clauses (King and Just 1991; Miyake et al 1994). Subsequent research used variations of this test, including a Listening Span test (presenting the stimulus sentences aurally), or a Speaking Span test (generating sentences from sets of single unconnected words, presented in increasing size). These studies also found correlations with verbal fluency and the ability to infer contextual meaning to learn new words more effectively (Daneman and Green 1986; Daneman 1991).

These two areas of research, into the phonological loop (or short-term storage) and loop/executive trade-off (or WM capacity), have formed much of the basis for using WM in L2 studies. Tests of the episodic buffer through, for example, prose recall, are not yet fully established in L1A (Baddeley 2000), but recent research (Fry 2002; Fehringer and Fry 2007) adopts a prose-recall metric from clinical psychological research, used to test for language impairment (the Adult Memory and Information Processing Battery, Coughlan and Hollows 1985), which is argued (Fry 2002) to implicate the episodic buffer. The prose-recall task involves listening to and then immediately repeating as accurately as possible a short narrative, with around ten schematic segments, lasting around 50 seconds, a task which the episodic buffer is suggested to facilitate (Baddeley 2000). Fry (2002) found significant correlation between native speakers' use of complex syntax in oral speech, including picture description and question-elicitation tasks (targeting optional adverbial, prepositional and subordinate clauses) with the capacity to immediately recall the Coughlan and Hollows narrative task accurately. This finding was matched in Fehringer and Fry's (2007) study of twenty proficient end-state bilinguals using similar tasks.

WM is thus seen to play a key role in L1A where novel information is to be processed (vocabulary acquisition) and complex or ambiguous reference is to be maintained in a task (reading, resolving ambiguity, prose recall). To recap, it is argued that native language processing is a combination of automatic and conscious processes, and that declarative and non-declarative knowledge interact in theoretically separate ways, but that they overlap extensively in language processing. In summary, WM as the "workspace" for conscious management of the necessary information and processes involved in a linguistic task is centrally involved. However, the role of WM in learning novel information and the interaction of WM and LTM in learning, especially in language, remains unclear (Jonides et al 2008).

One specific model for learning which does address the intersect between declarative and procedural memory is the Adaptive Control of Thought model developed by Anderson (1983) and further extended more recently (e.g. Anderson et al 2004). This model does specifically refer to memory in learning, especially in its earlier conceptions, such as ACT* (Anderson 1993), both in general learning theoretical terms and with reference to language. According to ACT*, knowledge which begins as declarative information can become proceduralised; procedural knowledge is learned by making inferences from new or already existing factual knowledge (or "instructions") using "production compilation mechanisms". These productions are developed by practice, so that initial reliance on instructions to inspect declarative knowledge slowly and consciously is converted into a fast, unconscious, automatic "set of productions for directly performing the task without declarative retrieval of the instructions." (Anderson et al 2004: 1046). This model does not specify how it applies to infants, whose procedural knowledge is stronger in their earliest years than their declarative knowledge (Baddeley 2007). Nor does it specify how great a role is played by WM. However, the model has been adapted by a number of studies in adult SLA as a helpful model in explaining the development of greater fluency and automatisation (e.g. Towell et al 1996; DeKeyser 1995, 2003; Segalowitz 2003).

Anderson's model draws attention to two commonly assumed features of the declarative-procedural or explicit/implicit dichotomy which have been central to discussions of mechanisms of native language use, and which are argued here to be crucial to understanding the question of what the L1 or L2 mind consists of: degree of automaticity and control.

2.6.6. Automaticity and control

Automatic language use is usually discussed, primarily, in models of speaking (Levelt 1989, 1999) or patterns found in sentence processing (Gibson 1998).

Natural speech, without conscious attention, is seen as a quintessentially automatic activity, dependent on implicit knowledge and procedural processing (Levelt 1989: 20-22). Levelt's seminal model of speech production proposed a tripartite model involving different procedural or automatic processors interacting with long term declarative knowledge. The three processing components were the conceptualiser for creating a preverbal message, the formulator for preparing the message as a grammatical parsed string, and the articulator for motor production in overt speech. His more recent model (1999) retains many of the same functions, but splits the operations slightly differently (see Figure 3 below) based on a bipartite division between a semantic/syntactic system and a phonological/phonetic system (1999: 86). Within the rhetorical/semantic/syntactic system, a preverbal conceptual message is generated, which requires choosing and ordering the relevant information from long term declarative knowledge, while keeping track of the current discourse (ibid: 87).

The conceptual message is translated into a grammatical structure through grammatical encoding. In order to do this, the formulator uses procedural or automatic processes of retrieving specific lexical information (lemmas) from semantic memory, or as Levelt expresses it, the "mental lexicon". The grammatical specifications of lemmas instigate syntactic building procedures or "frames" to build noun phrases, clauses and so on incrementally, through a process of selection and unification, which is affected by accessibility and salience (ibid: 99), and are stored in a temporary syntactic storage buffer (1989: 12). Levelt does not specifically refer to pre-formed schemas or idioms that may be stored as syntactically integrated chunks, but these could equally well be stored as lemma items.

The surface structure emerging from this first semantic/syntactic system then engages with the second phonological/phonetic system. The morpho-phonological codes attached to items in the mental lexicon operate on the preverbal message, allowing the structure to arrive at a "phonological score" as "parsed speech" (ibid: 87), and stored in an articulatory storage buffer. The parsed message is then given further prosodic shape, turning the structure finally into a phonetic or "articulatory" score", which is sent through the mechanics of articulation to produce overt speech.

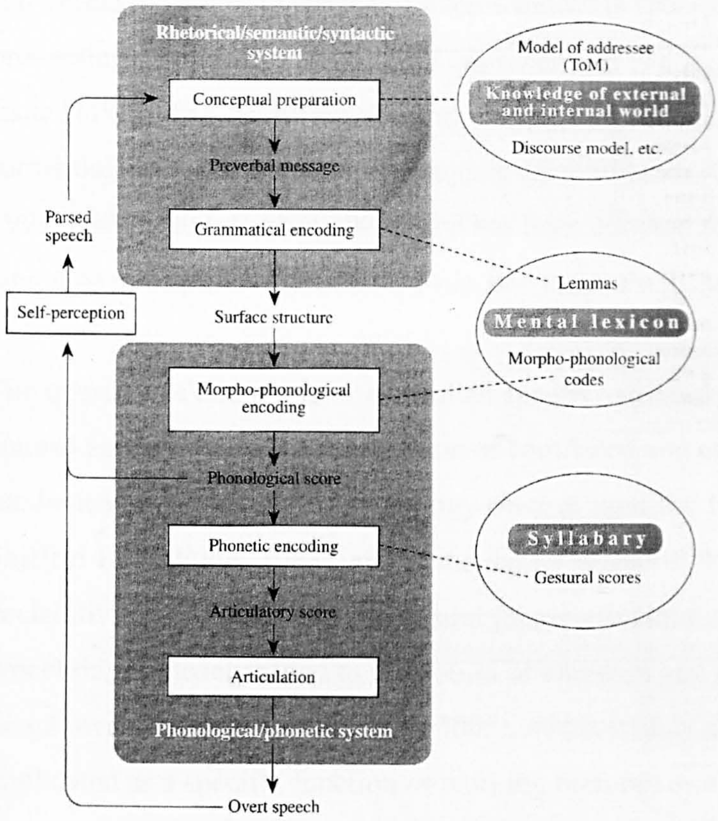


Figure 3: Levelt's (1999) blueprint of the speaker

Once articulated, covertly or overtly, this parsed speech can also be attended to or "self-monitored" via a feedback loop or speech comprehension system (ibid: 88), which returns slips or restarted speech to the preverbal message stage; monitoring does not always happen, since normal automatic speech does not necessarily or continuously require attention. The interim storage buffers suggested in Levelt's original model are part of "informationally encapsulated" automatic processing (Levelt 1989: 21), and thus separate from the notion of working memory, which according to Levelt, plays no role in automatic unattended language processing. However, conscious control of language, and thus WM, is required in the conceptualisation of the message, and in monitoring either internally or externally allowing for slips to be repaired, according to Levelt's model of production (ibid: 10).

Levelt's later (1999) model does not provide any different analysis to his (1989) model in terms of memory. It thus presents an interface between declarative lexical knowledge, which includes grammatical features attached to the word or lemmas, and procedural knowledge, which provides automatic procedures for encoding lemmas into full verbal messages in speech. Levelt's model is specified for speech production – the processing component for speech comprehension is not explored by Levelt as "not at issue" (1989: 13). However, the simplicity of the model has been widely accepted, and forms the basis of most bilingual models of mind/brain (Grosjean 1982, 2008; Paradis 1997, 2004, 2009; Fabbro 2002), and has been adopted for some models of second language perception or production (de Bot 1992; Poulisse 1997; Kormos 2006).

The question of automatic or controlled speech, outlined above, also raises the issues of control and attention. The separation of controlled and automatic processing has been a fundamental distinction in psychology since at least the 1970s (e.g Schneider and Shiffrin 1977; Fodor 1983), and is standardly linked to the distinction between declarative (controlled) and procedural (automatic) knowledge. Similarly, controlled processing is closely allied to the notion of attention and conscious awareness (Segalowitz and Frenkeil-Fishman 2005), and is widely agreed to be intrinsically implicated as a specific function of working memory as a "supervisory attentional system" (Norman and Shallice 1986; Baddeley et al 2009).

There are times when we use different, careful processes in speaking, when we monitor speech, in a controlled, non-automatic way, either if we are being very careful about precise terminology or complex grammatical structures, or when we rephrase or repair speech after we shift the semantics or syntax of what we want to say, or after making a slip of the tongue. In addition to using controlled attention to monitor online speech, we can also use controlled attention to deliberately recall and rehearse facts, including learned declarative rules about language. Just as those rules can be recalled in a controlled way and then used, so they can be thought about separately without necessarily being used: that is, analysed metalinguistically. Karmiloff-Smith (1986), Gombert (1992) and Bialystok (1994, 2002) have suggested models that reflect the development in children from using implicit procedural knowledge of language (from around 3 years onwards) to developing explicit knowledge about language. Skilful, accurate metalinguistic analysis depends on high levels of awareness and of control (Bialystok 1994), and is argued to require the attentional resources of working memory (Bialystok 2002).

In addition, Anderson's model, as shown above, assumes that declarative learning must be consciously attended. Early theories that information arrived in short term memory and by dint of sufficient attention or rehearsal became stored in long term memory (Atkinson and Shiffrin 1968) were subsequently revised (Craik and Lockhart 1972; Craik 2002). Craik's work was seminal in understanding that it is not how long an item is attended to that ensures its encoding and consolidation in memory, but how "deeply" or "shallowly" it is analysed. Deep processing refers to fully analysing information for "meaning, inference and implication" (Craik 2002: 308). Shallow processing refers to purely surface information of "surface form, colour loudness and brightness (ibid). Encoding is argued to be an automatic by-product of processing a stimulus, allowing for incidental or implicit learning (Reber 1967; Schacter 1987) where there is no specific intent or instruction to encode, but encoding is deemed to be more efficacious if attention is paid (Craik and Tulving 1975). These concepts of attention and automaticy are taken up in SLA by, amongst others, Schmidt's Noticing Hypothesis (1990), Clahsen and Felser's (2006) Shallow Processing Hypothesis, and work by DeKeyser (2003) and Segalowitz (1997, 2003) on automaticity.

Capturing data that truly distinguishes automatic implicit knowledge vs. explicit metalinguistic knowledge can be conceptually and methodologically difficult or controversial. Some types of automatic language use include, as mentioned at the end of section 5.4, sentence processing or parsing (e.g. Frazier and Clifton 1986; Gibson 1998), which probe when automaticity is impaired, through parsing breakdown in highly complex or potentially ambiguous sentences. However, I do not detail this research here, since much of the relevant key issues have been reviewed already in theories of linguistic representation and processing covered above. Generativist researchers (see also section 2.7.4.3 below) have also used offline or online grammaticality judgements as an appropriate way to test for automatic implicit intuitions of grammaticality, particularly for ungrammatical constructions (such as subjacency-constrained extraction) which are difficult to elicit. However, there have been many concerns raised about what such tasks are really testing (e.g. Birdsong 1989; Schutze 1996, 2005) in terms of accessing implicit or metalinguistic knowledge. The debate centres on whether such tasks are about testing automatic implicit intuitions (as assumed by generative researchers) or actually require high levels of control, awareness and metalinguistic knowledge (as argued above by Bialystok 1994, 2002).

Nevertheless, despite these caveats, the methodology is still widely used, especially in L2 (see further discussion in 2.7.4.3).

2.6.7. Conclusion of discussion of L1 acquisition

There have been many themes drawn out in this section on L1 linguistic representation, acquisition and use which form the basis of SLA theories. I have shown that there are both formal and processing constraints on complex structures such as wh-questions, especially when applied to long-distance extraction or subjacency constraints.

I have shown that there are different accounts for how children acquire language, and have drawn out what I believe are a relevant set of factors affecting language acquisition in the widest sense, i.e. child acquisition, and, by implication, SLA. Certain factors were deemed to indicate the nature of mature L1 capacities, such as automaticity, which arise from different types of memory, storage and retrieval that affect both knowledge and use, which I will argue also apply in SLA. Differing research paradigms make a unified account of implicit and explicit knowledge in LTM difficult to piece together with WM and how these both operate in language. In native language processing, only a limited role is assigned to conscious linguistic knowledge, and much of language use is deemed to be stable, efficient, procedurally encoded, implicit, unconscious and automatic. However, I have argued that WM is likely to play a key role in native language processing, especially for complex syntax. If so, then, by logical extension, WM can be argued to play an even more key role in L2 acquisition and use, particularly for instructed learners, where there is agreed in SLA research to be such wide variation.

I turn now to develop the constructs discussed above in the context of L1A to how they are researched in accounts of representation, acquisition and use in SLA.

2.7. Second Language Acquisition

Variability in L2 acquisition of the morphosyntax of English questions was chosen as the focus of this study as representing a number of key issues arising from the discussions above on L1 acquisition and processing.

Understanding the systematic variability observed in L2A, whether inter-learner or intra-learner, has long been the hallmark of SLA (Selinker 1972; Tarone 1988; R Ellis 1994). Different research strands focus on different explanations and contexts of variation.

Generative accounts of L2 acquisition have generally been taken to address variation only in terms of vertical or developmental changes in "competence". A key question in this context would be whether abstract linguistic knowledge of L2 features can be successfully acquired when these differ from the L1, especially for constraints argued to arise from the operation of universal implicit constraints on long-distance movement. There is the associated issue of whether asymmetries observed in different question forms in English would cause any marked difficulties for L2 learners of English. It has been noted that there are clear stages in L2 acquisition of morphosyntax. Short-movement lexical questions and embedded forms, followed by more complex long-distance extraction, are seen as later acquired (Lightbown and Spada 1993; Vainikka and Young-Scholten 1994; Pienemann 1998) and subject to variability especially in adult learners (Johnson and Newport 1989; Pienemann 1998). There has been debate in the literature over whether L2 learners can or cannot show sensitivity to L2 subjacency constraints, as contrasting evidence has been found in different studies (e.g. Schachter 1989; Johnson and Newport 1991; White and Juffs 1998). Asymmetries in processing have also been argued to play a role in observed differences between subject and object extraction (Schachter and Yip 1990; Juffs and Harrington 1995; White and Juffs 1998). Non-syntactic explanations have also been put forward to explain apparent difficulties in acquiring certain types of structures, which explain non-targetlikeness as difficulties mapping from underlying syntax to surface morphology (Prevost and White 2000).

In short, the mechanisms by which learners pass through the different stages of question acquisition are not yet fully understood. In addition, it is not clear why individuals vary so much both in rate and level of attainment. Furthermore, optionality between learners with the same language experience, and within the same learner at different times or between tasks, cannot fully be explained within the generative paradigm.

Variation has also been explained "horizontally" in terms of individual differences between learners, within a cognitive paradigm (e.g. Skehan 1998 and Robinson 2002). This strand of research assumes variation arises from individual differences in the interaction between "learning characteristics and learning contexts" (Robinson, 2002: 2). The difference between different learning contexts, e.g. the effects of instruction vs. immersion, however, is not always fully explained within this research paradigm, and intra-learner variability (or optionality), for example in showing both targetlike and nontargetlike morphosyntax, is also unclear.

This study thus focuses on adult instructed Chinese learners of English in an academic environment, in the face of evidence that complex morphosyntax, as found in English question formation, has been found to be problematic in academic discourse. Corpus-based analysis of over 1000 written academic texts from non-native students (Hinkel 2003) has shown that lack of complex structures including questions and subordinate clauses, default use of copula *be* (primarily in declarative structures) and difficulties marking tense appropriately, are among common indicators of non-native academic writing even after years of operating in English at an academic level. These findings are noteworthy given that question forms, especially simple short-movement, are a frequent and explicit focus of taught input from the earliest levels, certainly in Chinese and Taiwanese textbooks (e.g. Nani 2006), and both short and long movement questions are tested in intermediate examinations of English proficiency such as Cambridge First Certificate, Acklam 1996).

The combination of the prevalence of instructed input, combined with the traditional grammar-drill memorisation techniques common in Chinese and Taiwanese instruction (Gu 2003) could be argued to favour the development of explicit knowledge, which could be tapped as verbalisable metalinguistic awareness (Bialystok 2002) or of declarative or taught "rule knowledge" of the correct grammar (Mackey and Gass 2005: 51).

Working memory, facilitating the capacity to learn novel information and consciously manipulate knowledge, should therefore be closely implicated in the ability to learn and use such taught knowledge. Comparing acquisition of taught forms vs. implicit subjacency-constrained forms is assumed to give an insight into how the different types of innate, implicit and explicit knowledge overlap, compete or collaborate as required by the L2 user.

This assumption, arising from the review of the L1 literature, forms the basis for the research questions that drive this study. In order to provide the SLA background in more detail, I turn first to generative accounts of L2 acquisition, comparing selected theories to assess what they would predict for successful L2 acquisition of the structures targeted in this study (in section 2.7.1). Similarly to the sections above on L1, it is concluded that generative theories alone cannot fully account for individual differences in rate of acquisition or of transition through the stages of acquisition. Non-generative theories of SLA are then presented (in 2.7.2) that add to a greater understanding of variation in L2 acquisition, with a focus on issues affecting what kind of knowledge could be derived from instructed L2 acquisition, principally how the implicit/explicit distinction has been used in L2 research, the role of noticing and awareness, development of automaticity and control, and the effect of the input context.

In section 2.8, models of the L2 mind are presented, with particular focus on the MOGUL model of acquisition by processing, followed in section 8 by the presentation of two models that specifically differentiate explicit from implicit knowledge (Schwartz 1993 and Ullman 2004). Evaluation of these models lead to a discussion in section 2.9 of how WM is argued to play a role in SLA, and section 2.10 provides empirical evidence of previous WM studies in L2 acquisition and use. Finally in section 2.11, the motivation for this study and research questions are presented.

2.7.1. UG-constrained SLA

Generative approaches to SLA assume that the principles and processes of a language-specific faculty for L1A are implicated in acquiring an L2 (although they differ as to how far it applies and at what age). Evidence of implicit linguistic competence is found when L2 users display knowledge about linguistic structures they cannot have learned, but are deemed to have acquired incidentally from the input they have heard (primary linguistic data) - see White 1989, 2003; Hawkins 2001 for discussion of key issues and reviews of empirical data. Evidence that L2 users are able to behave like L1 users in processing their L2 fast, automatically and stably, particularly in speech, would indicate that L2 knowledge is procedural. Speed and accuracy in grammaticality judgements are also be argued to reveal access to implicit or procedural knowledge, although concerns have been raised (e.g. Birdsong 1989; Bialystok 1994; Sorace 2003 among others; see also section 5.6 above) that grammaticality judgement tasks may not necessarily tap implicit competence, but may in fact tap metalinguistic knowledge of the rules of taught grammatical structures and may also be confounded by other issues such as semantic plausibility and random guessing.

Different theories within this camp disagree both on how far UG applies to SLA, on the role of UG parameters transferred from the L1, and how input becomes intake (e.g. Eubank 1993; Vainikka and Young-Scholten 1994; Schwartz and Sprouse 1996; Hawkins and Chan 1997; Klein and Martohardjono 1999; Hawkins 2005; VanPatten 1996, 2005).

I turn first to review some of the key research that drives the assumption that UG remains accessible at any age, and that there is no fundamental difference between child and adult SLA. However, I focus in this study on adult SLA.

The evidence that suggests L2 acquisition is constrained in some way by UG is when L2 speakers show sensitivity to universal linguistic principles or language-specific parameter settings that are argued cannot be deducible from L2 input, from a speaker's L1, from instruction or from general learning or non-linguistic properties of mind (White 2003; Hawkins 2008). Much research was generated within the construct of the "Principles and Parameters" approach of generative grammar in the 1980s and 1990s. In this approach, acquisition is a question of resetting parameters which are different between the L1 and L2. Parsing mechanisms within the language module, which operate on exposure to ambient primary linguistic data, are sufficient to trigger change in implicit competence (Schwartz 1993; Vainikka and Young-Scholten 1998; Hawkins 2001). Klein and Martohardjono (1999) argue that, in a generative approach, grammars are reconfigured in an L2 through changes triggered when input is processed by the parser as intake (1999: 201) - "the parser, the mechanism which turns the input into intake, has a very direct role in grammar construction" (ibid). However, how the parser operates in turning input into intake remains unclear, as does the mechanisms of the language module interacting with the lexicon in L2.

UG principles, such as the Subjacency Principle and the Empty Category Principle, referred to in section 1.1, are argued to conform to a key assumption of generative acquisition, since they are seen as underdetermined in the input. Structures which violated these principles, being ungrammatical, would be neither taught in instructed settings, nor possible to learn from natural input (i.e. primary linguistic data). Sensitivity to these principles shown in target-like knowledge of the constraints on long-distance extraction by speakers of a language without overt movement, was argued to show evidence of UG affecting L2 acquisition. Studies during the 1980s and 1990s (e.g. Bley-Vroman et al 1988; Johnson and Newport 1989; Schachter and Yip 1990) found evidence that adults showed significantly different levels of accuracy in such judgements to children, leading some to hypothesise that UG was not available after puberty (see section 2.5.8 below).

However, other studies, such as White (1989), Martohardjono and Gair (1993), White and Genesee (1996), Li (1998), White and Juffs (1998) looked specifically at subjacency constraints and found evidence of native-like judgements, suggesting adult "access" was available. Li (1998) and Martohardjono (1998) also clearly identified a hierarchy of subjacency constraints between strong constraints on extraction from subject clauses, adjuncts and relative clauses, and weak, less unacceptable constraints such as *that*-trace violations and extraction from noun complements. White and Juffs (1998) tested two groups of Chinese adults, with and without immersion, and found that immersion made no significant difference in their accuracy levels on a grammaticality judgement task. White and Juffs' (1998) findings are very important in prompting the research questions for this study, and I therefore present their study in some detail here.

White and Juffs (1998) examined knowledge of subjacency in two groups of Chinese L1 learners of English, specifically comparing for the effect of immersion. Both groups (each with 16 participants) had been formally instructed in China from the age of around 11, but with no contact with native speakers of English before the age of around 16. The aim of the study was to see if the participants showed sensitivity to subjacency constraints, and whether living in the L2 country made a difference (1998: 115). The first group were postgraduates or professionals who had learnt English only at school, and then in intensive English classes at university (averaging around 10 hours per week). The second group had also learnt English at school in China, but had then moved to Canada for university study or teaching; they reported using English around 20 hours per week.

Participants took two tests, a timed grammaticality judgement test (with a binary choice of either grammatical or not) and an untimed pencil and paper question production task. They were tested on grammatical sentences involving NP + preposition (PP) phrases and extraction of subjects and objects from CPs (both finite and non-finite). This tested firstly if students had acquired wh-movement, and secondly if students had access to knowledge of subjacency constraints on wh-movement, so that they would find some way of rephrasing the question to avoid subjacency violations. Both tests were balanced for equal numbers of grammatical and ungrammatical forms, and for length of sentences.

White and Juffs found that both groups seemed to have acquired long-distance wh-movement, and observed subjacency constraints without a significant difference between native immersion and the Chinese environment. However, White and Juffs reported a subject-object asymmetry (as predicted in theoretical terms in section 2.1.4 above), as participants showed slower processing times on subject extraction than object extraction (also observed by Schachter and Yip 1990; Jordens 1991; Juffs and Harrington 1995). They propose that processing differences as well as theoretical factors underlie this asymmetry. They suggest that "L2 learners might achieve similar competence to native speakers, and yet take longer to access that competence" (1998: 127). This suggests that implicit linguistic competence in the L2 user is not as clearly associated with fast automatic procedural processing as standardly defined. Their conclusion has driven the interest of this research study into how best to investigate both representation and processing in the L2 user's mind.

Empirical evidence has also been found in studies from the 1970s onwards of a predictable pattern of development in L2 acquisition through different stages of morphosyntactic phenomena which followed a similar track to child L1 development, and which seemed to be independent of L1 transfer effects, since the pattern is common crosslinguistically for L2 learners, regardless of their original L1 (Dulay and Burt 1974; Bailey et al 1974). This evidence was also taken as support for the argument that generative principles and language acquisition mechanisms are common for L1 and L2

(Schwartz 1992), and underpins structure-building accounts of SLA (Vainikka and Young-Scholten 1996; Hawkins 2001).

The empirical evidence was further strengthened by extensive descriptive studies of developmental stages in L2A, outside the framework of generative SLA, which conformed to the predictions of Pienemann's (1998) Processability Theory (see further discussion in 7.1 below). The evidence thus points strongly, and independently of theoretical slant, that in L2 questions, there is a set order of acquisition, irrespective of L1, which follows an implicational hierarchy – learners cannot ask questions representative of a later stage, without being able to generate questions from earlier stages, except as holistic chunks (Myles 2004). This implicational hierarchy is shown below in Table 1.

This order is similar to that found in L1A (Stromswold 1989; Clark 2003), in which short formulaic chunks are acquired first, then wh-fronting without any verb, or without head movement, and head movement and auxiliary inversion acquired last (by around 3;0 years). Inverted copula is acquired before auxiliaries, and object *what* questions feature predominantly in both chunks and early constructed questions. An acquisition hierarchy has been found in children's wh-expressions, showing an argument/adjunct asymmetry, seen in developmental order as *where<what<why< who< when* (Clark 2003: 222).

Table 1: Order of acquisition of questions in L2 English (Pienemann, 1998)

Stage	Formation	Example
1	Rising intonation on words/ formulae	Four children? Do you know ...?
2	Rising intonation on clauses	The boys throw the shoes?
3	Double marking of verb; question word at front of clause without head movement	Is the picture has two planets on top? Where the little children are?
4	Copula fronting and inversion after wh-questions	Is there fish in the water? Where is the sun?
5	Head movement of auxiliaries, modals, "do"-support	Can you tell me? What is the boy eating?
6	Non-movement in embedded questions	Can you tell me what the date is today?

L1 transfer was not seen as critical in the face of these universal patterns of development in both L1A and L2A across different L2s. However, there remains an ongoing debate, even amongst those who argue for full adult access to UG, that acquisition of L2 must be affected by transfer of existing L1 settings to some extent (e.g. Schwartz 1999; Whong-Barr and Schwartz 2002; Sharwood Smith and Truscott 2006). I will now outline some key hypotheses that address this question, and their implications for acquisition of English wh-movement. These hypotheses are put forward to explain the role of transfer at the earliest stages of L2 acquisition (initial state), at developmental stages (interlanguage state) and the final stage (or end state).

One approach, based within the context of the maturational or structure-building hypothesis for L1 discussed above, is that L2 acquisition, similarly, starts with limited functional categories, and builds L2-specified functional categories "upwards". Vainikka and Young-Scholten, in their Minimal Trees Hypothesis (1994), and updated as "Organic Grammar" (2005) argue that only lexical categories, such as noun phrase (NP) and uninflected verb phrase (VP) are transferred, and not functional projections. A weaker approach is argued for by Bhatt and Hancin-Bhatt (2002) who find evidence of transfer of some functional categories, notably inflection or Tense Phrase (TP) but not CP. Hawkins' (2001) Modulated Structure-building hypothesis amalgamates elements of both "organic grammar" and full transfer. Learners build up their L2 syntactic trees from the bottom (VP) to the top (CP); L1 syntactic features are transferred and feature settings can eventually be reset, with sufficient input, but transfer only happens at the relevant stage in syntactic development (Hawkins 2001: 74).

Such approaches suggests a crucial interconnectivity between functional categories and UG constraints - as learners acquire L2 functional categories and L2 features, they will develop sensitivity to the constraints of relevant UG principles. Since there is full access, learners can acquire full native-like settings in their L2 functional competence, but would need to pass through the hierarchy of functional categories (VP before TP before CP), before they showed evidence of head movement, wh-movement and subjacency constraints.

These approaches argue that development through the hierarchy is based on the "interaction between the ambient, linguistic input and the learner's internal, innate linguistic mechanisms" (Young-Scholten et al 2005), and is triggered by overt morphology in the L2 input, following an implicational bottom-up hierarchy. In L2 acquisition of English and German, for example, the morphological triggers are argued to be free functional morphemes, rather than bound inflectional morphemes. The copula and modals trigger TP, while complementisers trigger CP (Vainikka and Young-Scholten 1998). These approaches suggest that L2 interlanguage grammars should follow the same stages of development as native grammars, and that ultimate convergence on the target grammar can be expected. Convergence, in this account, is taken as 60% accurate suppliance in obligatory contexts.

A number of concerns have been raised against this hypothesis, most notably from apparent evidence of TP and CP constructions in early L2 data, and also from evidence of L1 transfer in TP and CP constructions (Gavruseva and Lardiere 1996; Schwartz and Sprouse 1996; Haznedar 1997; Schwartz 1998). More crucially, issue has been taken with the concept that overt production of morphosyntactic elements (e.g. verb raising, tense marking, subordination) should be equated with the presence of the underlying functional category – in other words, the absence of overt morphosyntax does not need to mean the absence of the underlying abstract syntactic categories. I shall return to this point later, in discussing the Missing Surface Inflection Hypothesis.

A second, highly influential approach, is that of Schwarz and Sprouse's (1996) Full Transfer/Full Access Hypothesis which can be seen as a powerful argument covering both evidence of transfer and access: UG initially operates only through full transfer of all L1 settings, but can operate directly when the input does not conform to transferred L1 settings – in other words, syntactic development is "failure-driven" (Schwartz 1998: 147). This would in principle predict that speakers of languages without wh-movement could acquire full competence in L2 English wh-movement, and show sensitivity to constraints such as subjacency.

However, one of the key aspects of Full Transfer/Full Access (FT/FA) is that it makes no predictions as to how long L1 transfer would last. In principle, it is argued to affect early stages of acquisition only ("initial state"). It also does not guarantee ultimate convergence on the target L2, if the data needed to "force L2 restructuring" is either nonexistent or obscure (Schwartz 1998: 148).

FT/FA, and the other generative approaches discussed above, have a number of limitations, in that they do not account for variation between speakers of the same L1 at later stages of development who show different degrees of sensitivity to subjacency constraints, for example, when all other factors are equal (such as amount of exposure). Such variation would have to be explained away by some other factor, either that transfer, perhaps for processing reasons, would override access to differing degrees between individuals, or that non-linguistic factors such as motivation or attitude may influence individuals more than linguistic factors. In addition, it is unclear in FT/FA how to empirically test for evidence that the initial state does not consist of the L1 grammar, since even if there is data that shows non-L1/non-L2 influence, provided it is still UG-possible, the hypothesis stands. Nor is there any account for the evidence of developmental stages as discussed above. The FT/FA Hypothesis can therefore be argued to be so powerful as to be unfalsifiable (White 2003: 67).

A third approach highlighting the role of transfer was developed in Hawkins (2001) in his modulated structure-building hypothesis, building on the Failed Functional Features Hypothesis (Hawkins and Chan 1997), which has now evolved into the Representational Deficit Hypothesis (Hawkins and Liszka 2003). According to the Representational Deficit Hypothesis, for adult L2 learners, parameter resetting is particularly difficult, if not impossible, for features which are unspecified in the L1. Hawkins follows Adger (2003) in drawing a distinction between interpretable features (with semantic content) and uninterpretable features (without semantic content). Adger (2003) suggests that wh-movement is generated by an uninterpretable [*u*wh] feature which is strong in English and thus generates movement, but which is weak in Chinese. Hawkins argues that acquisition of uninterpretable features not instantiated in the L1 (such as Tense in Chinese) and differences in feature strength are subject to critical period constraints (Hawkins 2005: 128). Chinese learners of English are argued to be unable to acquire the strength of the [*u*wh] feature, and must thus interpret apparent wh-movement (including subjacency constraints) using licit L1 rules such as topic fronting (Hawkins and Chan 1997; Hawkins 2001). This argument is also made by Sorace (2005) and Tsimpli and Dimitrakopoulou (2007), suggesting that variability or optionality can occur most strongly for uninterpretable features, which are more prone to L1 transfer than interpretable features.

Hawkins' hypotheses that apparently convergent L2 grammars are driven by L1 settings seem to be undermined by counter evidence (White and Juffs 1998) of native-like success in judging subjacency violations and wh-movement constraints, where no L1 transfer can apply. This approach also does not provide an account for variation within and between L2 learners with the same L1 background and the same input. Nevertheless, the approach offers an explanation for potential asymmetry in wh-construction judgements, if there was evidence for systematic target-like judgements occurring in instances which could be interpreted as an L1-mediated strategy of topic fronting (such as argument extraction, allowed in Chinese), and non-target-like judgements for examples less amenable to L1 mediation (such as extraction from adjuncts and complex NPs, disallowed in Chinese).

In contrast to the earlier hypotheses which suggest that acquisition or non-acquisition of underlying syntactic features explains surface impairment in overt morphosyntax, another line of research, the Missing Surface Inflection Hypothesis (Lardiere 1998; Prevost and White 2000), suggests that the underlying syntactic tree may be fully present, with L2 features also successfully in place, but that surface impairment arises at the interface between syntax and morphology. In other words, a separationist model of grammar is proposed, in which the "output of syntactic computation is indirectly mapped via morphological (or phonological) module-specific translation procedures to actual phonological forms" (Lardiere 2008: 135). Variation arises as a result of difficulty not with syntactic competence but morphological competence (ibid: 111). In this account, morphological variation can be found after decades of exposure to the target language input, i.e. the end-state in L2 shows considerable overt fossilisation, but the divergence from the target is morphological rather than syntactic.

This hypothesis has been invoked to explain variation both at developmental stages and in ultimate attainment, and for children and for adults (e.g. Haznedar and Schwartz 1997; Lardiere 1998, 2007; Prévost and White 2000; Goad and White 2006). Lardiere's research into the language of Patty, a long-time speaker of English but who shows consistent morphological variability, has been particularly insightful into the possible dissociation between morphology and syntax. It is also a rare example of a much longer time frame (well over a decade) than many other longitudinal studies, and the findings are presented here in some detail.

Patty's data show that she has fully acquired wh-movement and head movement, and shows target-like judgements on subjacency, and other wh-related constraints. However, she shows evidence of variability in her use of question forms, showing both earlier stage-type questions - see Table 1 above (such as omission of copula, and do-support) - as well as fully fledged embedded questions.

Lardiere admits it is "not clear" (2007: 156) why Patty varies so much, and Lardiere does not provide evidence as to the amount of early-stage question forms compared to target-like forms; however, from the examples given in Lardiere (2007: 156-157), it seems as if the majority are in fact target-like. Lardiere (2007), overall, argues in favour of a degree of modularity of grammatical subsystems or domains (e.g. knowledge of phonology, syntax, semantics, lexicon, morphology), but that the "mapping" interface between some domains is more vulnerable to fossilisation than others. In Patty's case, free morphemes are more successfully produced than bound morphemes (e.g. irregular past compared to regular past). So she appears most vulnerable when mapping abstract syntactic features to the morphophonological spell-out of affixal inflection (ibid: 235). However, the development of clausal structure and knowledge of the features responsible for structural movement (e.g. verb raising for copula and modals, versus lexical verbs, use of do-support, wh-movement) have been largely successfully acquired (ibid: 236). Pervasive omissions "might be due" to non-language-specific processing mechanisms (ibid), and she concludes that "clearly both parts of the picture – a theory of UG-constrained representation and a processing theory – are needed to explain the data" (ibid).

In another approach, Goad et al (2003) argue that variability in L2 morphosyntax is related, at least in part, to L1 prosodic transfer. In consequence, there is a "discrepancy between learners' underlying knowledge of the L2 morphosyntax and their realization of overt morphology which must be mediated through nontarget-like (L1-based) prosodic representations" (ibid: 2). They note that in production (Bayley 1994; Lardiere 1998), Chinese speakers often have problems with final consonant clusters, but seem to exhibit more problems with regular past tense *–ed*, than with final consonants for irregular past (echoing Pinker's dual mechanism model comparing learned words vs. computed rules). Goad et al argue that aspect markers in Chinese are argued to be incorporated as "internal clitics" within a single Prosodic Word, whereas English tense markers which are bound morphemes (such as regular *–ed*, but not irregular past forms) are argued to adjoin to the lower Prosodic Word of the stem as part of a higher Prosodic Word. L1 transfer of incorporation, and failure to acquire L2 adjunction is suggested to as an additional factor (though not the only one) in explaining variation in production.

This hypothesis does not refer specifically to potential problems that Chinese L2 learners of English may have with wh-movement per se, but may affect variable or apparent failure to produce associated requisite head movement (subject-auxiliary inversion and do-support) if elements of tense and agreement marking are affected by prosodic transfer and differential effects of bound or free morphemes.

To conclude this section, a number of generative-based approaches to L2 acquisition have been discussed, which have focused on how successfully these approaches can explain variability in morphosyntax, especially for L2 question forms. Syntactic, morphological and phonological accounts have been proposed which explain some but not all the problems of variability and asymmetry that have been noted throughout the discussion of question formation in L1 and L2. These accounts will inform the methodological design of this research study, specifically testing the predictions of developmental models that can account for variability in intermediate stages of acquisition, such as Organic Grammar. In addition, in light of the possible impact on identifying successful L2 acquisition of difficulties at the syntax-morphology interface or of prosodic transfer, evidence from both written and oral data would help to identify how far non-targetlikeness may be attributable to morphological mapping difficulties, and will thus be taken into account in planning the data collection methodologies in this study.

One of the key points arising in the above analysis of generative approaches to SLA, raised by Lardiere (2007: 233) is how input (the stimuli of the target language) is processed by the L2 learner, a point echoed by Carroll (2001: 16) who notes that "what matters for language acquisition is how such stimuli are analysed". The failure to agree on what input consists of is evident in recent L2 acquisition discussions, such as Piske and Young-Scholten (2009), who provide a comprehensive but not always overlapping selection of different views on the nature of L2 input, including both generative and non-generative approaches.

The findings of studies such as White and Juffs (1998), discussed above, that implicit competence and processing differences need to be taken into account for a full picture of L2A underpins the view behind this research project that L2 language acquisition and processing involves a multi-strategic approach, and needs to take account of different types of knowledge, particularly in investigating instructed learners. One such approach can be found in Herschensohn's "coalition model" for SLA (1999). She argues "that the L2 learner uses a coalition of resources including a UG template, L1 transfer, primary linguistic data and "instructional bootstrapping" (ibid: 220). The coalition thus integrates a "general cognitive awareness of grammatical principles" as well as "substantive universals instantiated by L1". This knowledge is "presumably located outside the language module in the knowledge base" (ibid: 184-85). I discuss this issue in more detail in section 2.8, where I review models of mind in L2, to address the question how different types of knowledge may represented and used in the L2 user's mind.

I turn now to theories of SLA which assume no role for UG in adult L2 acquisition. These theories and models specifically focus on adult L2A as part of general learning and cognition, based on L1 cognitive research (as discussed in section 2.4.2 and 3.2 above) and therefore address squarely the difficulties in explaining individual differences and variability in L2A in terms of processing, memory and individual cognitive differences in interaction with input. These theories provide a logical context to address the role of memory and working memory in L2 development or transition, which generative approaches to date have largely side-stepped.

2.7.2. Non-UG theories of adult SLA

Theories which see no role for UG in adult L2 acquisition are largely based either on the premise that UG is available for children but not for adults, or on the cognitive premise that UG plays no role in language acquisition at all.

Much of the non-UG based adult SLA research has been dominated by Bley-Vroman's (1990) Fundamental Difference Hypothesis (FDH). This argues that whether or not UG plays a role in L1, there is too much evidence of dissimilarity of outcome in L2, especially adult L2 learning, for UG to be accessible in adult SLA. Therefore this position assumes that adult SLA is fundamentally different from both L1 acquisition and child L2 acquisition. There is an assumption that age of exposure to the L2 is crucial, in that successful targetlike acquisition is limited by a so-called "Critical Period" (Lenneberg 1967; see also Harley and Wang 1997; Herschensohn 2007). According to some researchers, adult L2 is driven by general learning mechanisms, but mediated by or in competition from UG instantiated in the L1, and will therefore not show native-like levels of accuracy (e.g. Bley-Vroman 1990, 2009; Clahsen and Muysken 1986; Schachter 1989; Felix 1985).

In a very recent revision of the FDH (2009), Bley-Vroman acknowledges the need to review the nature of any remaining fundamental differences between L1 and L2 acquisition, in light of the changes in UG-theory discussed above, and calls for research questions that address in both L1 and L2 the nature of shallow versus deep processing and other psycholinguistic learning mechanisms that he now sees as valid in explaining phenomena in both L1 and L2 (such as subjacency constraints).

As mentioned above, Schachter (1989), and Johnson and Newport (1989, 1991) failed to find similar levels of accuracy in post-puberty L2 learners compared to native speakers. They concluded that SLA was fundamentally different to L1 acquisition, and, for adults, is thus primarily driven by general processing constraints. Johnson and Newport's studies have had a marked contribution in subsequent research, and are important to this study in their focus on L1 Chinese and wh-movement, and I present their studies in some detail here.

Johnson and Newport (1989, 1991) tested 44 native Chinese or Korean adult speakers of English, both of which L1s lack wh-movement. All participants had lived in the US for at least 3 years prior to testing, and were divided between early arrivals (arriving in US before the age of 15) and later arrivals (after the age of 17). Johnson and Newport compiled a grammaticality judgement task of 276 items testing twelve different elements of English morphosyntax, including verbal inflection (past tense, present progressive, 3rd person singular), verbal subcategorisation, particle movement, pronominalisation, plurals, auxiliaries, yes-no questions and wh-questions, and word order. They found in their study robust evidence that late learners were significantly less target-like than early learners – the correlation between age and overall performance for the early group was highly significant ($r= -0.87$, $p<.01$), while there was no such finding for the late group ($r=-0.16$, $p>.05$). Correlations between age and accuracy on rule type were also found, with the highest correlations found for past tense, pronouns and plurals (all over .70, $p<.01$). Correlations for yes-no questions and wh-questions were at or below .50, but again were significant ($p<.01$). These findings were strengthened by further analysis of subjacency tokens (Johnson and Newport 1991), which were judged by late learners at or below chance, and drove their conclusion that UG was unavailable to late learners.

A number of studies on processing times for wh-constructions (Juffs and Harrington 1995; Marinis et al 2005; Felser and Roberts 2007) also find that adult L2 learners do not show evidence of recognising traces in wh-constructions in the same way that native speakers do, especially in terms of differences between subject and object extraction. Clahsen and Felser (2006) suggest in their "Shallow Processing Hypothesis" that adult L2 learning mechanisms rely more on shallow parsing driven by lexical cues than on abstract syntactic elements of phrase structure such as traces, and conclude that adult L2A is fundamentally different from L1A. However, this hypothesis could be tied to the suggestion raised in White and Juffs (1998) that processing rather than representational differences separate adult L2 users from native speakers, and I conclude that differences in processing may not provide overwhelming evidence that the underlying source of implicit linguistic knowledge is different.

2.7.3. Cognitive theories of SLA

Extending the assumptions of the Fundamental Difference Hypothesis (FDH), many studies of adult L2 learners based within a wholly functional-cognitive paradigm assume that language acquisition is driven by general learning mechanisms, whether child or adult, L1 or L2 (e.g.DeKeyser 2000, 2003; N. Ellis 1994, 2001; McDonald 2006; Robinson 1996, 2001, 2003). This paradigm thus goes beyond the FDH in rejecting any role for generative nativist constraints even in childhood. There is also increasing interest in applying the connectionist model to L2 acquisition (N. Ellis 2005;

Williams 2005), to explain L2 acquisition in terms of shifting weights or strengths of activation levels.

Cognitive-based accounts of both L1A and L2A assume that the same mechanisms, such as frequency, salience and schema-matching, which operate in response to the input, work throughout the lifespan, and therefore apply similarly to L2A (N.Ellis 1994, 1996, 2005; Williams 1999). In addition, language use has been argued to be key to development, both in interaction with interlocuters (Long 1996; Gass 1997) and in production of spoken output which drives and is driven by increasing linguistic knowledge (Swain 1985).

There are three key themes within this paradigm, which I have drawn on in some of my assumptions behind this study. The first theme is drawn from cognitive psychological theories of implicit or explicit knowledge and learning (Reber 1967; Paradis 1997, 2004, 2009; N. Ellis 1994; Hulstijn 2005), the role of implicit and explicit instruction, and "focus on form or formS" (Doughty 2001, 2003; R. Ellis 1994; R. Ellis et al 2008; Robinson 2002; Norris and Ortega 2000; Sanz and Morgan Short 2005), and the centrality within this paradigm of the role of attention and noticing (Schmidt 1990, 2001). The second theme focuses on characteristics of successful acquisition, such as automaticity and fluency (DeKeyser 2001; Segalowitz 2003), which also draw directly on cognitive psycholinguistic constructs, such as Anderson's ACT* model (1983, 1996). The third theme places a greater emphasis on context of learning and different types of language use, such as study abroad (Freed 1996; Freed et al 2004; Sunderman and Kroll 2009), and interaction (Long 1996; Gass 1997; Mackey 1999). A final theme within this paradigm directly concerns the role of individual differences such as WM in SLA (Miyake and Friedman 1998; Robinson 2001), and how this might interact with studies of aptitude (Skehan 1998, 2002). Miyake and Friedman's study specifically looked into how far WM represents elements of language aptitude (Skehan 1998, 2002), which the authors specified as analytic language capacity, memory ability and phonetic coding ability (Miyake and Friedman 1998: 340). The broader strand of aptitude research falls outside the scope of this study; however, Miyake and Friedman's claim that WM is "the key to variation" in second language acquisition is a challenge that this study goes some way to address, and the theme of WM in SLA is separately discussed in later sections (2.10, 2.11).

The cognitive research paradigm has been criticised as sometimes adopting a "broadbrush" approach to linguistic phenomena (Mitchell and Myles 2004: 192). Its typical focus on a Western or Anglophone classroom or laboratory environment constrains the comprehensiveness and applicability of many of the specific findings (ibid). However, the available research provides valuable evidence as to how learners process input, and what contexts or methods for processing input can lead to more efficient processing, greater accuracy, and fluency. The themes identified above thus provide the context for my hypothesis that WM plays a central role in SLA.

2.7.4. Themes within cognitive theories of SLA

2.7.4.1. Implicit/explicitness

The first theme regards the standard explicit/implicit distinction drawn by N. Ellis (1994, 2001, 2005) and R. Ellis (1994, 2005) and Hulstijn (2002, 2005), as it is argued to relate to knowledge, instruction and learning in SLA.

N. Ellis argues (1994, 2001, 2005) that the two forms of knowledge and learning are distinct and dissociable but they cooperate in language acquisition, which is usage-driven. Explicit learning of language occurs through conscious efforts to negotiate meaning (Long 1996; Gass 1997). N. Ellis assumes that much of adult L2 language is learned implicitly, based on associations, chunks and constructions drawn from the input through frequency and usage, through which linguistic prototypes and categories emerge (2005: 306). However, many aspects of a second language are "unlearnable" from implicit processes alone (ibid), and he claims that both implicit and explicit learning occurs in SLA in co-operation through a dynamic interface – learning can be kickstarted during conscious processing or noticing of explicit input, which the learner registers patterns and constructions, which are then "integrated into the system by implicit learning" (2005: 305). Explicit knowledge can also guide the conscious creation of novel linguistic utterances through processes of analogy from existing memory stores, including formulas and declarative pedagogical grammar rules, whose subsequent usage promotes implicit learning and proceduralization. Learning also occurs through output (Swain 1985), even "flawed output" (N. Ellis 2005: 305), since this can prompt focused feedback by way of recasts that present learners with further linguistic data for analysis (ibid). WM is implied to play a central role in explicit learning as the means of facilitating noticing and providing workspace for analysis (Ellis and Sinclair 1996).

However, research into implicit/explicit learning is affected by a range of definitions and assumptions as to what the terms mean. R. Ellis (1994, 2004, 2005, 2008) and R. Ellis et al (2009) have consistently argued for the importance of distinguishing implicit and explicit learning, and implicit and explicit knowledge. Ellis (2008: 6-7) defines the terms using standard psychological terms referred to earlier for L1 (e.g. section 2.6), as follows:

> "*Implicit knowledge* is intuitive, procedural, systematically variable, automatic and thus available for use in fluent, unplanned language use. It is not verbalizable. According to some theorists it is only learnable before learners reach a critical age (e.g. puberty).
>
> *Explicit knowledge* is conscious, declarative, anomalous and inconsistent (i.e. it takes the form of 'fuzzy' rules inconsistently applied), and is only accessible through controlled processing in planned language use. It is verbalizable, in which case it entails semi-technical or technical metalanguage. Like any type of factual knowledge, it is potentially learnable at any age."

R. Ellis thus maintains the standard psychological distinction between these two types of knowledge (Paradis 2004, 2009). By contrast, others (particularly DeKeyser (1998, 2003) follow Anderson's (1993) skill-based theory, and thus argue that declarative knowledge can evolve into procedural knowledge through practice. There is also some discussion in the psychology literature that there is less distinction between the two types of knowledge than are commonly presented, and should be seen in terms of a continuum, rather than a dichotomy (Dienes and Perner 1999). Ellis acknowledges that linguistic evidence supports some type of continuum, though he argues this may appear more in relation to different degrees of consciousness and ability to articulate knowledge of language in use, than theoretical representations of that knowledge itself. Thus, he concludes (2008) that despite separation at the level of representation, there is wide acceptance that explicit and implicit sources of knowledge interact at the level of performance (see also Schwartz's discussion of explicit learning vs implicit acquisition in section 2.9.1 below).

Hulstijn (2002, 2005) elegantly draws out the concepts and issues at stake in discussing implicit and explicit theories of SLA, in particular the difficulty of definitions and the lack of consensus over possible definitions (2005). He focuses on the learning mechanisms that draw out regularities underlying the input, whereby implicit learning is input processing without awareness of such regularities, but explicit learning is "input processing with the conscious intention to find out whether the input information contains regularities and, if so, to work out the concepts and rules with which these regularities can be captured" (ibid: 131).

Hulstijn argues that processing input could take place either implicitly or explicitly since all of these learning skills could, arguably, be applied either consciously or unconsciously. He warns against mixing the process of learning (*how*) with the object of learning (*what*) (ibid: 133). He concludes that learning is dependent on both implicit and explicit mechanisms, and argues that noticing is key, certainly for explicit learning, as outlined in the following section, in the sense that "input" needs some kind of conscious attention to become part of the learner's internal processed "intake". His emphasis on noticing reflects a central assumption with the cognitive SLA paradigm that noticing is crucial to L2 learning (Schmidt 1990, 2001; Robinson 2001; Doughty 2001; N. Ellis 2005).

However, just as definitions of explicitness and implicitness remain debated, so there is conflicting evidence of the effectiveness of explicit compared to implicit instruction as a way of driving development, at least in classroom learning (see overviews by Norris and Ortega 2000, and by Sanz and Morgan-Short 2005). DeKeyser (1995) testing learning of an artificial mini-language, found evidence that explicit rule presentation led to better performance in an immediate post-test and on a delayed posttest. Robinson (1997) found that learners given explicit rule presentation outperformed other groups who were given implicit, incidental or enhanced input, when measured using grammaticality judgements for structures representing "hard" and "easy" pedagogical rules (pseudo-clefts of location, e.g. *Where Mary and John live is in Chicago not in New York.* – hard, and subject-auxiliary inversion with adverbial location phrases, e.g.

Into the house ran John – easy). However, other studies, such as Rosa and O'Neill (1999), did not find any beneficial effect for explicit rule presentation. Such findings highlight the danger of assuming too quickly that instruction should be the same as learning (Carroll 2001; Sanz and Morgan-Short 2005) – just because teachers provide copious explicit or implicit instruction, it is not sufficient to guarantee that input becomes intake. Another limitation of many such studies comparing explicit and implicit instruction is that they usually only test for effects of "acquisition" over a limited period (typically a few weeks – DeKeyser's (1995) posttest was twelve weeks later), and White's (1989) study on adverb placement famously found no permanent effect.

Other approaches have sought to overcome this input-to-intake gap by manipulating the input through explicit instruction to process the input differently (such as VanPatten's Processing Instruction approach – which I discuss further in section 2.8.1 below), or implicitly through input flood (Trahey and White 1993) or input enhancement (Sharwood Smith 1991). As with explicit instruction, findings are less than clear as to the effect of such approaches, and how widely applied they can be for different constructions (Sanz and Morgan-Short 2005), and again, there is often a limitation that posttest effects are measured in relatively short timescales. These findings beg the question whether what is being tested is akin to medium term recall or some kind of task practice effect, rather than real lasting restructuring (intake). Sanz and Morgan-Short (2005) also argue against adopting too simplistic a dichotomy in such research, suggesting that a more gradient continuum is needed, distinguishing more clearly the impact of differences in input, complexity of the target linguistic structure, task requirements and practice effects (2005: 249).

Doughty (2001, 2003) discusses the constructs of explicit/implicit knowledge, noticing and working memory in some detail within the context of instructed learning. She suggests adults learn via general learning mechanisms, as a result of "developmental sharpening" (2003: 299), specifically analytic problem-solving mechanisms, and predictive comprehension. She reviews evidence from a wide number of previous studies and concludes that it is unclear whether explicit L2 instruction is efficient in producing long lasting acquisition. She also believes that implicit learning is possible, in fact more effective in language acquisition terms than metalinguistic instruction, and believes the two approaches can overlap to drive acquisition, in that declarative knowledge is a by-product of practice during implicit learning (2003: 295). She has investigated the capacity of learning methodologies which include overt focus on form, within the context of communicative activities, to encourage adult L2 learning to return to the kind of "discovery mode" that she presumes children use (2003: 299). Instruction using input enhancement via Focus on Form and other means, will, she claims, assist adult L2 learners to return to implicit response to input (mainly surface cues). Focus on Form is argued to be beneficial for recalcitrant problems, e.g. for L2 users who have been long-immersed and are fully able to comprehend language, but continue to display residual L1-based interference (2001: 206).

N. Ellis and Sinclair (1996), N. Ellis (2005) and Doughty (2001) connect Schmidt's noticing hypothesis (Schmidt 1990, 2001) with the construct of WM in SLA. They emphasise the need in SLA to "notice the gap", to trigger cognitive comparison (Doughty 2001: 225). Doughty claims that WM "serves as the interface between everything we know and everything we perceive or do" (2001: 227). She specifically analyses how the construct of short term selective attention may work for focus on form without interrupting the flow of fluent language use, drawing on models of WM such as Cowan (1999) which have a variable length of temporary activation (up to 20-30 seconds), to suggest that there may be a "window for provision of focus on form" (Doughty 2001: 226). Ellis (2005) argues that WM provides the cognitive workspace for attentional selection and perceptual integration for storing and using novel information (2005: 305).

2.7.4.2. Noticing and attention

Noticing, in the sense of conscious awareness of information, is held to be a central component of learning, closely linked to Baars' theory of the Global Workspace (Baars 1997) and has become a key element in cognitive theories of SLA (Schmidt 1990, 2001). Learning is argued to be mediated by "attended processing" (Schmidt 2001: 29) or noticing. Noticing or attention is defined by Schmidt in a relatively restricted sense, overlapping but not isomorphic with awareness (which is seen as a separate construct). Noticing, in his terms, means "detection plus rehearsal in short-term memory" (2001: 5).

Schmidt argues that attention assists "nearly every aspect of second and foreign language learning" (2001: 6) both for learner-internal factors (including current L2 knowledge and processing ability) and learner-external factors (including input, context, instruction). Attention is argued to have limited capacity, be selective and voluntarily controlled, activates conscious awareness, and is necessary to ensure that temporary or new information can be "registered" long enough to be processed or stored in memory. Attention is argued both to facilitate information beyond an initial pre-conscious automatic processing stage, and to inhibit non-relevant information in order that more relevant information can be processed.

However, there is ongoing debate over the exact nature of the different aspects of noticing, attention and registering, and how far conscious awareness is relevant to any or all of these aspects (Schmidt 2001: 18; see also Sharwood Smith 2008). Schmidt suggests that a difference should be made between non-conscious registration and conscious noticing within focused attention (2001: 20). In terms of L2A, the construct of noticing, as defined here, would mean that the learner makes most progress in learning various forms of knowledge (such as lexical meaning) when they pay direct attention to it. Within a functional approach to L2A, where the usual triggers of learning are argued to be frequency, salience and meaningfulness, L2 input could often fail to trigger learning by being "infrequent, non-salient and communicatively redundant" (2001: 23). Intentionally focused attention on specific elements to be learned through raising awareness, such as explicit instruction, input enhancement, focus on forms, can therefore prove helpful (as VanPatten, Sharwood Smith, Doughty amongst others have argued).

Schmidt admits that the question of how far attention may or may not be necessary for learning remains unanswered (as Hulstijn 2005 noted), and acknowledges the growing body of evidence from priming studies of learning of unattended items (Schmidt 2001: 28), as well as from naturalistic learning of complex or non-salient syntax without obvious attention or instruction (2001: 31). He does not specifically address how the construct of conscious attention relates to "parsing" mechanisms (2001: 32). However, given his above definition that noticing involves short-term memory, then it is logical that WM capacity constrains noticing capacity (Robinson 2001), although I note that

WM is only one of various factors that may constrain noticing, such as prior knowledge, and cognitive style (Skehan 1998).

For the purposes of this study, I argue that detection without attention is analogous to implicit learning, whereas detection with attention (or rehearsal) is analogous to explicit learning. In this discussion of implicitness and explicitness in SLA, it is clear that individual differences such as WM capacity should affect acquisition in terms of capacity to notice explicit input, through instruction and, and if adults are less able than children to employ implicit learning mechanisms, then WM should be vital to SLA.

2.7.4.3. Automaticity and control

Turning now to features of L2 processing, and their similarity to those discussed in relation to L1, I pointed out (in section 2.6.6) that automaticity was taken to be a central defining feature of L1 processing, and in discussions of bilingualism, with either matched or variable proficiency (1982; Cook 1997; Segalowitz and Frenkeil-Fishman 2005; Kormos 2006), the degree of automaticity and speed of processing is also assumed to be a key indicator of fully formed implicit competence or proficiency in language use.

A huge body of research investigates how far bilingual processing is the same or different to monolingual processing (Grosjean 1982; de Bot 1992; Paradis 1994, 2004; Cook 1997; Poulisse 1997; Fabbro 2002; Bialystok 2002; Kroll and de Groot 2005). Many of the research questions apply equally to SLA as to bilingualism, but it is a highly vexed question how far SLA and bilingualism overlap, especially in terms of advanced or "near-native" L2 attainment (see, e.g., de Groot and Kroll 1997; Hyltenstam and Abrahamsson 2003). Since this study focuses on learners who do not appear to be at an end state, I assume that such learners should be studied in the context of theories of L2 transition and development, so I do not provide details here of research based more firmly in the field of what can be called "balanced" bilinguals (Grosjean 1982).

There has been much research which, either specifically or in effect, investigates SLA in terms of increasing automaticity or implicit knowledge through comprehension and production. Many studies have looked at sentence processing tasks (e.g. Juffs and

Harrington 1995; Juffs 2001, 2004; see also Harrington 2001 for an extensive and detailed overview comparing L1 and L2), or speech comprehension and production (e.g. Dechert 1980; Segalowitz and Segalowitz 1993; Kormos 2006). Models that suggest a fundamental difference between L1 and L2 are often based precisely on evidence of different, slower, shallower processing found in L2 than in L1 (e.g. Cook 1997; Clahsen and Felser 2006). There is also some discussion over whether the identification of proficiency in terms of developing towards nativelike speed, accuracy and fluency is helpful. Speed and accuracy are not guaranteed measures of implicit knowledge, since target-likeness in taught forms can also derive from correct encoding of explicit knowledge, which can simply be "speeded-up" (Segalowitz 2003). Equally, nontarget-like forms can be misencoded either as implicit or explicit knowledge, and quickly or invariably retrieved.

The generativist paradigm (see section 2.6.5) has standardly measured implicit competence as fast, invariable awareness of ungrammaticality on constructions that do not exist in L1, are underdetermined in natural input and have not been taught. Offline or online grammaticality judgements have been widely used as an appropriate way to test implicit intuitions of grammaticality, particularly for ungrammatical constructions (such as subjacency-constrained extraction) which are difficult to elicit. However, there have been many concerns raised about what such tasks are really testing (e.g. Birdsong 1989; Schutze 1996, 2005; Sorace 2003) in terms of accessing implicit or metalinguistic knowledge. The debate centres on whether such tasks are about testing automatic implicit intuitions (as assumed by generative researchers) or actually require high levels of control, awareness and metalinguistic knowledge (Bialystok 1994, 2002). Paradis (2004) doubts how far grammaticality judgements are either purely tests of explicit knowledge or true tests of implicit competence. He views metalinguistic knowledge as part of general cognition but stresses that having sufficient metalinguistic knowledge to judge a structure does not provide the implicit competence to produce that structure online. He concludes (2004: 58) that an L2 user may successfully complete a grammaticality judgement task using either explicit metalinguistic awareness or automatic implicit competence, thus making it hard to truly assess the source of knowledge. However, despite these caveats, the methodology is still widely used, usually in conjunction with other tasks to capture a wider set of data to increase validity, and will be adopted in this study (see chapter 3).

However, explanations of how learners become more automatic are widely found. The most common assumption in the majority of cognitive accounts reviewed here is that proficiency derives either from implicit learning, or by explicit learning which becomes, through practice, increasingly automatic. DeKeyser (1997) found evidence of automatisation through practice which he argued supported Anderson's ACT* theory of increasing proficiency (see section 2.6.5 above). Explicit assimilation, then practice leads to automaticity, which is measured by reduced reaction time, reduced error rate and decreased interference from simultaneous tasks. DeKeyser noted that automaticity was not necessarily task-general – improvements in comprehension through comprehension practice activities did not transfer to good production, and vice versa.

In terms of developing L2 speech, Kormos (2006) provides a detailed analysis of second language speech perception and production, from a usage-based perspective, and argues that WM capacity plays a key role in developing more fluent automatic L2 speech (see also Kormos and Safar 2008). She stresses that speaking in L2 creates very high attentional demands on learners who are still in transition from less fluency to greater fluency, which is conditioned by WM's responsibility for regulating attention in cognitive processing (Baddeley 2003). She holds that non-fluent L2 speakers must hold already processed verbal information in memory in while planning or linguistically encoding the next segment of their utterance. Levelt would argue that this kind of temporary storage does not require conscious attention in normal fluent speech, and therefore is contained in an encapsulated "buffer", not in working memory. But, in Kormos' view of L2 speech, this temporary storage must be contained in working memory, since conscious control is needed to hold the pieces of the linguistic string together while they are being retrieved serially from long-term memory, or self-monitored for accuracy and comprehensibility. Similarly, in reading or listening (Kormos and Safar 2008), WM is implicated in the ongoing comprehension process, in which the L2 learner has to hold already processed bits of the text in memory as well as read or listen to the next part simultaneously, otherwise they will not be able to understand the text as a whole (2008: 267). It is possible, following Jackendoff (2002), or Caplan and Waters (1996) - see section 2.5.1 and 2.6.3 above, that such working memory stores are domain-specific for language only, but there does not yet appear to be sufficient evidence to clarify such a hypothesis, and Kormos sees Baddeley's model as most suitable for her analysis of conscious control in L2 speech.

Looking at L2 development in terms of greater automaticity, Segalowitz (2000, 2003) raises a note of caution of how much research still needs to be done, not least since the construct itself remains not yet fully "operationally defined" (2003: 403), e.g. comparing instance-based retrieval versus proceduralisation through analogy. In work based on ACT*-based skill development, Segalowitz stresses the need to see automaticity as not simply "speeding up" (2003: 387). He discusses the possibility of assessing changes in processing, e.g. by using coefficient of variability (RT SD/RT mean) rather than just RT itself. He notes that qualitative changes in speeding up may not always equate to better or more accurate language processing, e.g. speech, where non-target like forms become "easy" to retrieve, thus creating "fossilisation" – but he does not discuss precisely what the processing mechanisms of retrieval are in this scenario. However, he acknowledges that speeding up can lead to better language use, e.g. in reading, where quick matching of word to meaning, or to schemas stored in memory, can free up more resources for comprehending complex sentence structure, or whatever is required. Where knowledge that appears automatic in one mode (e.g. word recognition) does not transfer to another mode (e.g. word production), this lack of transfer argues against instance-based learning.

2.7.4.4. Context of input

One of the questions arising from the implicit/explicit instruction dichotomy focuses on how different types of exposure may affect the capacity of input to become intake. This is a question that transcends the generative/cognitivist divide, although the way input is discussed differs to some extent. As has been shown in earlier sections (e.g. section 2.4.1), input (whether in terms of naturalistic primary linguistic data or as either explicit or implicit input) is seen as necessary to trigger change or transition through different stages of development. However, input alone is not sufficient to explain such change, and has not really been addressed within generative approaches to SLA (although see e.g. Flege and Liu 2001; Moyer 2004; Rothman and Iverson 2007: Piske and Young-Scholten 2009). In particular, it remains open as to whether immersion or study abroad programmes provide a markedly different type of exposure or input for instructed learners than their classroom-based learning.

Evidence from some of the first intensive scrutinies of the effects of more intense exposure compared to what is commonly offered in foreign-language instructed classrooms is found in Canadian studies on immersion-type programmes in the 1980s. The general consensus was that immersion learners could demonstrate native-like competence in listening comprehension and reading skills, but generally fell behind native speakers in their productive language skills such as writing and speaking (Genesee 1987; Harley and Swain 1984; Swain 1985). This was seen to be due to the large amount of non-native speaker input that the immersion students received, since the only native speaker typically was the teacher.

Flege and Liu (2001) assessed different groups of L2 speakers of English, comparing students to non-academic workers resident in the US for either less than 3 years or more than 3 years, to see if the kind of engagement with L2 academic life promoted acquisition over workplace immersion where rich native L2 input may be restricted. They investigated a number of phonological and grammatical measures of targetlikeness comparing students to non-students, who had had either more or less than three years' residence in the US, and found that long-stay non-students scored significantly lower than the long-stay students. However, the short-stay students also scored slightly higher than the non-students, regardless of length of stay. They assumed this was due to the difference in type of input, assuming that students engaged in constant use of the L2 in a highly demanding academic environment promoted acquisition.

However, the effect of immersion per se compared to more common study-abroad contexts remained unclear. More recent studies investigated different types of context with mixed findings. Isabelli (2004) identified that for a small group of learners of Spanish, their accuracy in tense, aspect, and agreement features improved after study abroad. Howard (2006) also found that morphological accuracy in French verbal inflections improved for advanced learners who had spent time studying abroad.

However, others found little or no benefit for study abroad on linguistic development, especially for acquisition of implicit syntactic constraints argued to be facilitated by naturalistic input. Rothman and Iverson (2007) tested thirty intermediate learners of

Spanish before and after study abroad for evidence of acquisition of constraints on null and overt subject pronoun co-reference in Spanish (the Overt Pronoun Constraint), which is argued to show evidence of access to Universal Grammar, and is not formally taught as explicit metalinguistic learned knowledge. They found evidence that two thirds of the instructed learners were able to show native-like knowledge of the constraint prior to immersion, but that five months of immersion providing "exposure to abundant amounts of naturalistic input" (Rothman and Iverson 2007: 288) was not sufficient to trigger acquisition of the constraint among the ten other participants.

Other studies also found no effect for study abroad on morphosyntactic or lexical development (Collentine 2004; O'Brien et al. 2006), and it is more commonly observed that typical study abroad programmes have significant effect on oral fluency and communicative competence in a broader sense, rather than specific measures of grammatical accuracy or complexity (Freed 1995; Freed et al 2004).

A separate theory that has been argued to be relevant when considering both the impact of naturalistic immersion or study abroad, as well as the importance of noticing, is that language development can be triggered by interaction (Long 1996; Gass 1997). Research has investigated whether certain types of interaction, including recasts and feedback, can trigger restructuring since they focus learners on noticing specific elements within the input, in a type of orally enhanced input (see overview by Gass, Mackey and Pica 1998). In a study specifically looking at question forms, Mackey (1999) explored the ways in which learners' grammatical proficiency in question forms changed after interaction with native speakers. Mackey found that interaction facilitated grammatical development through the higher stages of Pienemann's hierarchy of question forms shown earlier. Thirty-four lower-intermediate adult learners at language schools in Australia were tested on oral production of question forms during one week's treatment with differing degrees of passive or active interaction, and were then tested again in a delayed posttest five weeks later; development was greatest for participants actively interacting with native speakers. Admittedly, the native speakers had been trained in how to direct their interactions to focus on question forms rather than meaning, and it is possible that the kind of more communicative responses that are usual in natural interchanges in an immersion environment may not necessarily drive development in the same way. However, I infer that the requirements of interaction in a naturalistic environment would have some linguistic effect on previously instructed learners with little exposure to native-speaker interactions (Gu 2003).

One of the assumptions raised at the start of this chapter thus addresses the question of investigating how far instructed learners develop, when they have previously had little exposure to naturalistic input, and are then immersed in a richer target-language setting. It has been investigated whether WM would facilitate development through greater noticing of the richer input environment, as discussed in earlier sections on noticing and awareness. Mackey et al (2002) specifically address the question of whether WM capacity is implicated in noticing feedback and linguistic development in question formation, but I will review this study in more detail in section 2.11, with other studies using WM in SLA.

Due to limitations of study design, this research study does not specifically control for strategies of noticing or detailed inspection of exposure to input either in instructed or immersion settings. However, by comparing patterns of linguistic development across a single L1 group of immersed L2 users, when prior instructional background and L2 immersion experience (i.e. all postgraduate students) is controlled for homogeneity as far as possible, this study is aimed to provide some insight into how change in type of exposure interacts with individual differences such as WM.

To conclude, the discussion in this section of key issues under generative and cognitive accounts of SLA provides a basis to investigate the challenge set by Gregg (1996) for SLA to provide both property and transition theories of what the nature of underlying linguistic representation might be, how it can vary between individuals, and how other factors such as L1 transfer, and type of exposure, affect the acquisition process. The issues remain greatly debated, and there appears to be as yet little consensus between the two paradigms. However, it is clear that both generative and non-generative researchers agree that native language is characterised by unconscious or implicit linguistic knowledge that can be automatically and rapidly accessed. The different sides principally differ only in whether this kind of knowledge is in a separate language faculty or not.

O'Grady, in a recent commentary discussing the debate between generative nativist accounts versus emergentist accounts of SLA in a special edition of *Lingua* (2009) stresses the point that much of the functional-cognitive assumptions of language acquisition do not deny the role of innate and implicit learning, merely the generativist claim that there is a language-specific device containing language-specific computational principles. He calls for a view of the "radical middle", which differs from connectionism in its conceptualisation of hierarchic symbolic representations, but which differs also from generativism in not positing inborn categories or principles that are specifically grammatical (syntactic) in nature. In the same issue, R. Hawkins, working within a generative paradigm, asks, "What kinds of L2 knowledge result from inductive learning from linguistic experience? Do non-linguistic principles, such as O'Grady's linear processor, play any role in a 'linguistic nativist' theory of SLA? Should some kinds of knowledge, that up to now have been assumed to derive from UG, be eliminated from the language faculty because they derive from more general principles of mind or can be explained by experience?" (Hawkins 2008: 619).

Such calls make it clear that what is needed is greater consensus as to how linguistic and cognitive factors interface, and what role memory, and specifically WM, plays in such a possible interface. I turn now to models of the L2 mind, which attempt to dovetail some of these factors together.

2.8. Models of the L2 mind

A number of theories and models are put forward for L2 in which processing and memory need to be taken into account. Models which focus more on processing include, most notably, VanPatten (1996, 2004, 2005) and Pienemann (1998, 2003). Other models which focus on possible architectures of the mind, in terms of acquisition through processing, are Carroll (2001), Truscott and Sharwood Smith (2004) and Sharwood Smith and Truscott (ms). These models all, to some degree, highlight how automatic implicit knowledge can be attained in the L2, and assume differing roles for attention and working memory.

2.8.1. Models of processing

VanPatten, working within a generative framework, has specialised in addressing morphosyntactic phenomena that appear to be resistant to L2 input from a psycholinguistic perspective. His model of Input Processing is directly aimed at informing pedagogical intervention in order to aid acquisition, and places working memory central to this model.

VanPatten does not discuss acquisition per se (along the lines of feature-setting discussed in section 6.1-6.4), but looks at what may impede acquisition in the context of how learners process the input primarily for meaning. He emphasises the notion that what drives acquisition is not the total input per se but intake (Corder 1967), which VanPatten calls the "subset of filtered input that the learner actually processes and holds in working memory during on-line comprehension" (2005: 271). VanPatten specifies that what is held in working memory is the data which is processed for meaning (not simply "noticed" from the input - 2005: 277), so he distinguishes IP from form-based "explicit" or "awareness-based" strategies outlined above in discussing definitions of explicitness or implicitness. Thus real-time processing constraints are vital to his concept of how acquisition is driven and what can impede successful acquisition. Acquisition is argued to be impeded in forms that are redundant, even if meaningful, and the emphasis on lexical meaning can drive learners to misappropriate form-meaning relations, such as encoding *do* and *did* as question particles marking yes-no questions, rather than as an auxiliary verb surfacing in questions and negatives to show tense and agreement information (2005: 278). However, he does not refer in detail to a specific construct of working memory or long-term memory, as far as I have been able to establish.

Myles (1995, 2004) has also addressed the question of how fluency and grammatical accuracy interact, by investigating the interface between learned chunks, derived from explicit presentation and repeating drilling in early instructed learning, and the development of implicit competence. She noticed that early learners use chunks as a strategy for producing apparently more fluent language than the underlying syntactic knowledge base supports (Myles 2004). Evidence of chunks being used rather than analysed syntactic rules could lead to misproductions such as *Comment la fille je m'appelle* (literally "How the girl I am called?", intended meaning "What is the girl called?"). She concluded that such early chunks could eventually, as competence developed, facilitate later reanalysis of the syntax within the chunk. This hypothesis supports the argument that learners may well rely on both explicit and implicit sources of knowledge at once, and that the different sources provide what can be most easily processed when communication of meaning is prioritised, as Van Patten and others argue.

Pienemann's Processability Theory (1998, 2005) is one of the most influential models that specifically addresses the question of how processing drives acquisition. His theory is that L2 acquisition is implicit, just as in L1, in that linguistic knowledge is symbolic but is derived from linear word order and sentence form. The theory itself is designed to be neutral on the underlying source of implicit knowledge, and the hierarchy of predicted stages of development in the acquisition of question forms, has already been presented in section 2.7 above. [6] Pienemann himself operates the presumptions of processability within the framework of Lexical Functional Grammar (following Bresnan 1982 - see section 2.3.1). The model is based on a hierarchical development of "processing procedures" (ibid: 14), starting with the lexicon up through the "lexical category, phrasal procedures, to matrix-subordinate clause processing" (ibid: 13). The implicational nature of the hierarchy is that a higher level of processing cannot precede a lower level, and levels cannot be jumped. Pienemann states that the task of acquiring a second language "includes the acquisition of the procedural skills needed for the processing of the language" therefore tying speech processing and production more integrally into the acquisition process (2005: 2).

The theory was developed specifically to explain paths of L2 development (ibid: 36), and is careful to distinguish between processability components (acting as procedural processes) and automatic linguistic knowledge itself. However, Pienemann does not address the question of how the learner may utilise potentially overlapping sources of knowledge (e.g. learned chunks or metalinguistic knowledge of constructions or rules), and his emergence criterion of acquisition (three examples of a target form found on different lexical verbs or nominal phrase) could be seen as much less exacting than other definitions of acquisition (ranging from 90% of suppliance in obligatory contexts, taken from the L1 literature on child acquisition to 60% suppliance, argued by Vainikka and Young-Scholten 1996). The emergence critierion potentially obscures issues of variability, when a learner interchangeably uses forms from either stage, but it is not

[6]T he predictive assumptions of processability that underpin Pienemann's hierarchy are echoed in the predictions of generative structure-building approaches based on morphosyntactic triggers, that yield a very similar hierarchy – Young-Scholten et al 2005)

clear what is constraining such variability. Pienemann also does not specifically discuss the questions of individual optionality, inter-learner variability, the role of awareness or of working memory as a means of filtering input.

I turn now to two major theoretical attempts to meet the challenge raised by Gregg (2001: 179): "If we want to content ourselves with general learning mechanisms like attention and memory, we are going to have to show how they can interact with input to produce L2 knowledge, i.e., a grammar".

2.8.2. Models of acquisition

2.8.2.1. Carroll – Autonomous Induction Theory

Carroll's Autonomous Induction Theory (2001) builds on Jackendoff's Parallel Architecture (see section 2.5.1 above). She presents a modular approach to language acquisition, in which UG provides the theory of linguistic knowledge within an autonomous language module. This theory of representation is combined with a theory of learning driven by processes which use both implicit (inductive) and conscious (metalinguistic) processing.

Carroll's theory of SLA focuses on four elements:

i. A theory of linguistic knowledge (a theory of mental grammars)
ii. A theory of knowledge restructuring (how the representations of a mental grammar can change in principle and, equally importantly, how they cannot)
iii. A theory of linguistic processing showing how input gets into the system from the speech signal (bottom-up) or from the conceptual system (top-down) thereby creating a learning problem
iv. A theory of learning which shows how novel information (not brought into the system from outside the grammar by the parsers) can be created to resolve the learning problem (2001: 39)

She takes UG as the first element required - a theory of linguistic knowledge - after a discussion of a wide range of empirical evidence from supporters and opponents of UG. However, she rejects UG theories, such as Principles and Parameters, as insufficient for the second element to explain how knowledge is restructured for adult SLA, concluding that Principles and Parameters "is at best a metaphor" and one "that has outlived its usefulness" (112).

To explain the second, third and fourth elements of her theory, she adapts two models from psycholinguistics and psychology, namely, Jackendoff's model of autonomous modules of language and Holland et al's (1986) induction model. Her adaptation of Jackendoff's (1987) Representational Modularity model has been somewhat overtaken by Jackendoff's later reworkings of this model (see 2.5.1 above) but the essential components remain applicable.

As outlined above, three autonomous representational systems are linked by correspondence modules (integrative processors), which operate bottom-up or top-down between each system. Thus, for example in speech parsing, an external stimulus (such as an acoustic soundwave) is translated upwards to phonological format, then into syntactic format, then to conceptual format (for comprehension), and encoded representations are stored in long term memory.

However, Carroll makes a crucial difference in defining input, not always drawn in other models, between what is heard and what leads to learning. She separates out stimuli (what is heard) and intake (what can be parsed) from input (what is processed). She defines intake is "transduced stimuli": intermediate representations to the speech parsers (10). The parsers are defined as mechanisms designed to encode the signal in various representational formats, which can encode grammatical distinctions and are tuned to be "language specific". Input is what is actually being processed - "any mental representation which is analysed by a processor" (14), whether intake or second-order representations drawn from long-term memory.

She sees SLA as different to child acquisition. In child acquisition, the logical problem is laid out in terms of children acquiring a first language where the representational systems for conceptual representations, phonological and syntactic representations must all develop, if not from scratch, then from a level less developed than an adult's. By comparison, adult learners of a second language by definition already have a representational system fully developed in the L1. Hence SLA does not suffer a logical problem of acquisition but an empirical problem of development. However, she does not clarify how far the L1 hinders or affects L2 acquisition and how the adult learner's processing mechanisms successfully "choose" an L2 representation rather than an L1.

Carroll is careful in her use of terms in discussing what is being acquired. In the context of a lengthy discussion of the implications of the distinction between competence and learned knowledge (Schwartz 1993), she chooses to adopt the term I-language as neutral term to describe "all aspects of an individual's knowledge of language" (24), which is unspecific as to how and where the knowledge is represented, and does not clarify how variability in use occurs. The construct thus includes implicit psychogrammar (encoded information which is directly relevant to the parsing and production systems), and also explicit metalinguistic information encoded in the conceptual system (24).

Her view is that acquisition of linguistic knowledge (or i-learning) is failure-driven, similar to other generative researchers, particularly Schwartz (e.g. 1993). She argues that language-specific implicit induction processes (condition/action or "IF/THEN" type automatic rules) drive i-learning or re-structuring of i-language, which are triggered by failure to parse the analysed input, either through detectable errors or a mismatch in the closeness of fit between input and currently activated representations (168).

However, unlike Schwartz, she dismisses a possible difference between competence (acquisition) and learned knowledge. She argues that there is no "empirical evidence for the distinction" (258) between the two types of knowledge, but rather argues that any kind of input, whether classroom or naturalistic, can provide the necessary "linguistic stimuli" to drive acquisition (ibid). This seems to hinge on her interpretation of Schwartz that input (as in external acoustic information) is processed strictly bottom-up, and that information from the conceptual system cannot interact with grammatical information.

However, there are a number of difficulties in applying her approach to the constructs of implicitness and explicitness outlined in earlier sections.

Carroll, it seems, adopts Jackendoff's view that the language module is itself relatively autonomous from general cognition, so that language restructuring does not derive from "a great undifferentiated general theory of learning" (393). The processes she describes such as input, restructuring, encoding, are all assumed to be implicit. However, she seems to equate "conceptual representations" (26) with propositional or declarative knowledge, which, I have argued in section 2.6 above, is consciously learned; perhaps in some sense the conceptual system overlaps with the autonomous processes she discusses. This assumption is supported by Carroll's view that an over-strict definition of modularity, such as Fodor (1983), has "no empirical basis" in SLA (Carroll 2001: 285)

Although Carroll does not elaborate on this issue, it is possible that the overlap inferred above reflects work done elsewhere on how lexical knowledge interfaces between an autonomous language module and general cognitive knowledge of word meanings (Whong-Barr 2005). Emonds (2000) distinguishes two parts of the Lexicon which straddle a theoretical divide between language module and general cognition (along the lines of Smith and Tsimpli 1995, discussed in section 2.5 above). There is, first, the Dictionary that lies at the interface of general cognition/memory stores (presumably declarative, although this is not specified), which contains real-world meanings of words, and, second, the Syntacticon, within the language module, which contains linguistic specifications and features driving syntax and morphology.[7] However, I have not found clear evidence in Carroll or Jackendoff (2002) to ensure this is how Carroll conceives the use of declarative lexical knowledge in her theory.

Additional clarity would help in the discussion of the role of parsing. Carroll defines parsing or speech production as encoding linguistic information in working memory in real time, which involves bringing unconscious knowledge to conscious working memory. She also talks about the capacity to analyse knowledge in terms of "representational redescription", where autonomous representations, e.g. from parsing or production systems, correspond with representations in the conceptual system, where it can be put to use for conscious reflection, performing metalinguistic tasks (26) but

[7] It also remains debated to what extent morphology itself should be seen as separate or not (Emonds 2000), and if so, whether it should be seen as part of the lexicon, part of a syntax module, or in a separate autonomous module, as is argued by Lardiere (2009).

does not specify how that correspondence is made in any psychological construct of memory or processing.

The overlap between what is unconscious and what is conscious is not clear. Carroll also suggests inductive learning can involve both implicit and explicit learning (127) – but makes nothing more of this hypothesis. She draws on a wide range of literature to assess whether intervention, such as feedback and correction, can help. She concludes that feedback and correction can only play a limited role, and primarily when there is awareness that feedback has corrective intent (394).

An important point arising from Carroll's assumption that learning is failure-driven, is that i-learning stops (fossilisation) when no detectable errors will be registered by the parsing system, though the learner may still be making systematic errors in production (169).Therefore comprehensibility of the signal/input by the L2 learner underpins lack of further change, not errors in production. However, she does not make much of this, and also does not provide a principled explanation for variability between target-likeness and non-targetlikeness.

So, to conclude, in trying to match her analysis of language acquisition to other psychological constructs, it seems that that acquisition or i-learning is largely unconscious (implicit) but non-modular; instructed learning can also drive i-learning by providing linguistic stimuli, so the input context is not conceptually or empirically different. Knowledge about language, or conscious metalinguistic analysis, can also play a role, but this role is not specified.

2.8.2.2. MOGUL

Jackendoff's model underpins another model which attempts to integrate UG representations with a theory of real-time processing. The Modular Online Growth and Use of Language model (Truscott and Sharwood Smith 2004; Sharwood Smith and Truscott, unpublished draft ms, cited with permission) develops the construct of Acquisition by Processing; the model specifically addresses the issue of different sources of knowledge, and stresses the importance of L1 transfer to explain variability, adopting non-generative connectionist theories of levels of varying and competing strengths of activation.

Truscott and Sharwood Smith (2004 - henceforth TSS) and Sharwood Smith and Truscott (ms – henceforth SST) suggest that there is a single processing system for language both for comprehension and production, and unified for first and second language use. They suggest, following Jackendoff's model, that there are parallel language-specific syntactic and phonological systems within the module and a conceptual system of "primitives". TSS suggest that the conceptual system (CS) is located outside the language-specific module, partly to explain the explicit nature of much of learnt lexical knowledge (2004: 4). They adopt the word "lexicon" to refer to the whole operation of the language module and associated non-module stores. Thus they conceive the lexicon as three long-term stores, phonological system (PS), syntactic system (SS) and the conceptual system (CS), which they call "sub-lexicons" (2004: 3). Lexical processes are the processes which occur at the interfaces between the systems both within and at the edge of the language-specific module. These structures are illustrated in the figure below (from Sharwood Smith, with permission), showing how information stores and processors are connected bidirectionally and also connected between each other by interface processors. The conceptual store interfaces with the inner levels of the language module proper, but is not strictly linguistic (ibid).

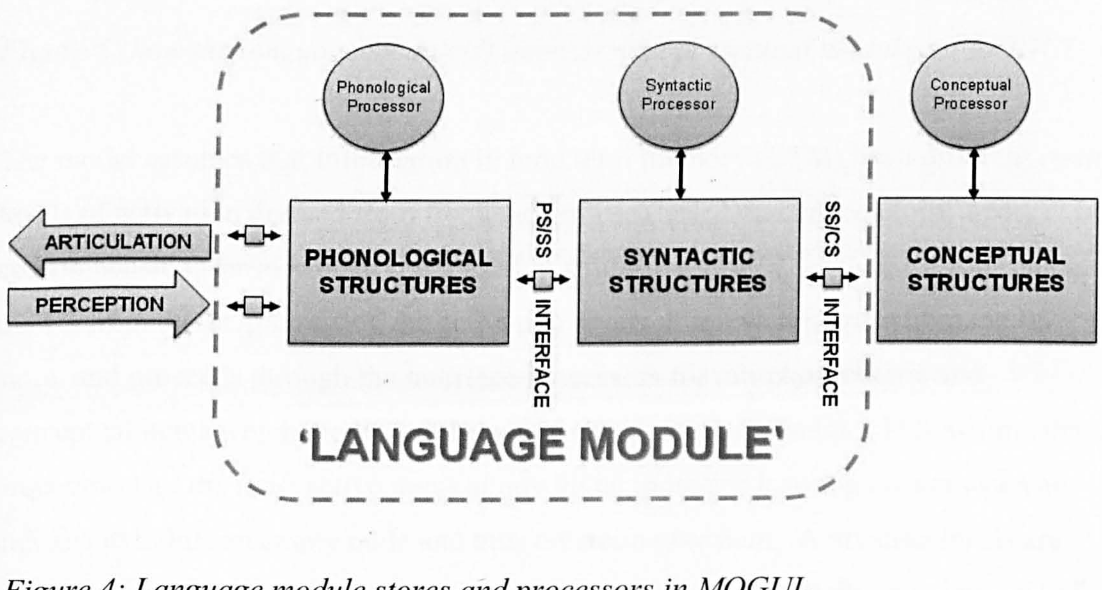


Figure 4: Language module stores and processors in MOGUL

The inner language modules (SS and PS) and the conceptual module (CS) are connected to the non-linguistic perception system through a "composite blackboard" of extralinguistic modules for perceptual processing and representations of sensory input, called perceptual output structures (POpS) (SST, ms: 94) – see the figure below (adapted from SST, ms: 100) for a more detailed illustration of the POpS. These POpS are accessible to consciousness and serve as the basis for higher level processing. It is the interaction and synthesis of all these processors inside and outside the language module which contribute to the "message" (the output or the input).

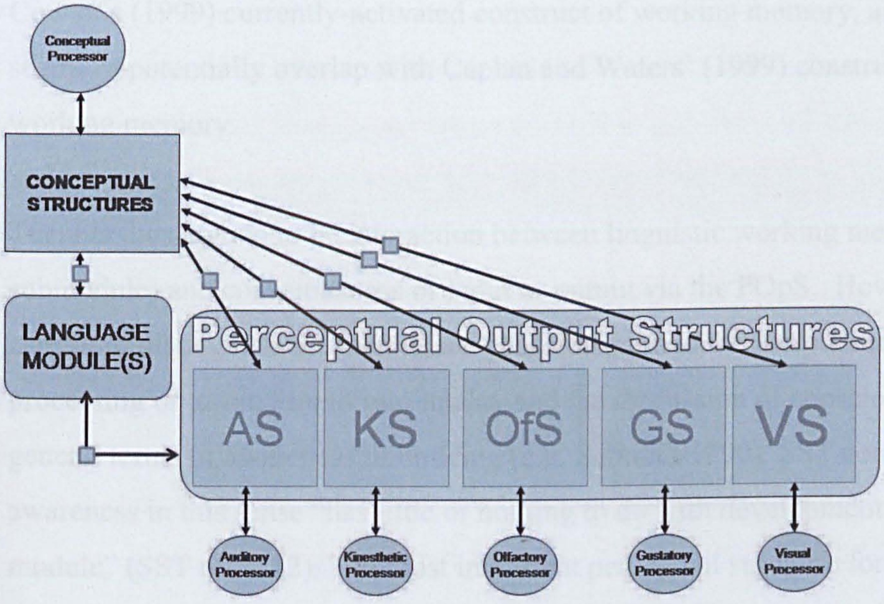


Figure 5: how the language module(s) interact with perceptual modules in MOGUL

The model assumes that information in long term memory (LTM) have different resting levels of activation derived from frequency of use (based on connectionist theory). In comprehension, sensory input (here used in a more generic sense than Carroll) enters the PS from the POpS, raising the activation levels of relevant items within the PS store, and proceeds through the interface processors for morphosyntactic and conceptual items which are linked through a common typing index. TSS assume that a processor uses the most active items at any given moment; learning occurs when an indexed item hits an empty node and thus creates a new item. Activation levels are raised as the incremental processing operation continues, potentially in competition for the best "legitimate" interpretation (2004: 5); when the processor rejects an item as no longer needed activation levels return to resting. Priming provides an example of "lingering effects" showing raised activation levels – if a subject hears a sentence in a passive form, they are more likely to contemplate a possible passive form in a potentially ambiguous following sentence (2004: 6).

Like Jackendoff, the model also involves working memory explicitly, but in order to fit the model, WM is seen, not as a separate "blackboard" or alternative construct to LTM, but as the upper activated layer of LTM, "a transient pattern of activation of elements within long-term memory stores" (SST, ms: 39). The emphasis, as in Jackendoff's model of language-module-specific working memory, is explicitly connected to Cowan's (1999) currently-activated construct of working memory, and additionally seems to potentially overlap with Caplan and Waters' (1999) construct of verbal working memory.

There is thus argued to be interaction between linguistic working memory (within each submodule) and consciousness of input or output via the POpS. However, SST are careful to distinguish between raised consciousness in the narrow sense of driving processing or turning input into intake, and the discussion of consciousness in more general terms of awareness or noticing (e.g. Schmidt 1990). SST conclude that awareness in this sense "has little or nothing to do with development of the language module" (SST ms: 212). The most important perceptual structure for language is deemed to be the auditory structure, although the visual structure is also important. These two can be equated in some part to Baddeley's phonological loop and visual-spatial sketchpad; the additional POpS for other sensory input (e.g. olfactory and kinaesthetic) are an improvement on Baddeley's model, which as it stands has no construct for other types of sensory input other than auditory or visual. Baddeley's most recent model (2000), to include the episodic buffer notes that the episodic buffer's role is to "combine information from the loop, the sketchpad, long-term memory, or indeed from perceptual input, into a coherent episode" (Baddeley 2007: 148). This function can be argued, as SST do, to mirror their construct of the function of POpS (SST, ms: 101).

However, SST stress that the auditory structure is not intrinsically linguistic or limited to phonological input (and so can also interpret music or birdsong); they highlight how the structure would thus deal with novel sounds such as clicks which may be phonological in one language system, say Xhosa, but not recognised as such in another language, say, English (ms: 96). Thus the auditory structure (or auditory loop) can be argued to be as effective a construct as the standard phonological loop for learning novel linguistic information.

SST argue that acquisition of novel information is form-meaning-based and usage-driven. For acquisition of functional categories, or syntax, SST argue that the syntactic store (SS) contains a pre-existing set of variable features and possible values (derived from UG), which must be co-indexed with a set of semantic primitives (such as the setting for T to include future time) in CS. An input with a novel SS feature or functional category, such as the Spanish setting for [future] feature for an English learner, is handled when SS can create a new item for the novel category in the learner's specific set of activated SS structures, copied into CS to grasp the meaning. Increases in the resting level of SS and CS items come from their continuing use by the processors in constructing representations. There are no failure-driven learning mechanisms as such (as argued by Carroll 2001), since acquisition represents changes in levels of activation, or "the lingering effects of processing" (ms: 88).

Transfer, as discussed in section 2.7.1 within UG models above, does not therefore apply in the same way. Instead, TSS adopt a structure-building account for an L2, based on the notion of "no transfer/full access" (2004: 15), since their view of L2 is that L2 knowledge (as it grows) exists side-by-side with L1 knowledge, but at differing levels of item activation of the required features. For example, in L2 English, V is acquired first, largely by lexical learning; then the inflectional category I (or T) is acquired when the copula must be parsed from the input. Once T is established, then the features [strong] or [weak] can be derived from the input, which drive the position of verbs, adverbs, negation and so on (with an argument that perhaps a default setting is weak). L1 features in the L1 lexicon (e.g. strong T) and L2 features in the L2 lexicon (e.g. weak T) compete, and "transfer" is the effect of L1 activation levels entering into L2 formulation and comprehension because their higher activation levels "win the competition" (ms: 69). Only sufficient input and output producing very high activation levels can drive the resetting of the L1 value.

MOGUL offers much potential to the challenge of finding a coherent transition theory for SLA. MOGUL's account of competition between strong versus weak levels of activation dovetails especially well with theories such as the Missing Surface Inflection Hypothesis (Prevost and White 2000), and is helpful in accounting for variability or optionality between L1 and L2 settings (see also Truscott 2006). However, a greater discussion of empirical evidence would need to establish, for example, how missing inflections arise from overriding L1 feature strengths. It also needs to be established whether MOGUL can account for evidence of variability in L2 data which do not conform to L1 feature settings; also for data reflecting UG-compatible settings that are not part of L1 or L2. It may be that novel nodes can be activated by some kind of default setting even when these are [weak] (as inferred by SST, ms: 70). MOGUL may also provide an account for differences between comprehension and production, where judgements on specific structures appear native-like in comprehension, but are non-nativelike in production, if it is assumed that online production requires more processing to draw on weak L2 values than comprehension, but SST do not, as far as I can see, explore this in their current account of MOGUL.

MOGUL also potentially fits with other competing system hypotheses (Felix 1985; Rothman 2008). Felix (1985) is an early account of variability in L2 acquisition closely associated with Bley-Vroman's Fundamental Difference Hypothesis (1990), that adult L2A is based on general cognitive systems rather than an innate language-specific cognitive system. However, Felix specifically assumes that any difficulties in attaining targetlike competence in an L2 result from competition between two autonomous cognitive systems – the adult's existing L1 innate competence, and the general problem-solving cognition system which is "fundamentally inadequate for the purpose of language acquisition" (1985: 51). In its original conception by Felix (1985), there was, as he admits, "not enough known about "the cognitive principles underlying problem-solving behaviour" (ibid: 68), and his insight about the competitive nature of differing sources of knowledge was not widely taken up particularly among generative SLA research.

However, in a recent revisiting of the concept of competing systems, Rothman (2008) renews interest in Felix's original insight, incorporating more recent developments of the relevant psychological and linguistic concepts, and additional empirical data which have been established in the intervening two decades.

Rothman focuses on persistent nontarget ("target-deviant", 2008: 76) morphology in tense and aspect for instructed L2 learners as a symptom of competing systems. He acknowledges that this focus overlaps with ideas and data underpinning the Missing Surface Inflection Hypothesis, but suggests that the MSIH alone "may not be able to explain" the disparity between targetlike competence found on the structures being investigated, especially in the face of lingering variability in highly advanced learners (ibid: 83). He argues strongly that pedagogical metalinguistic knowledge forms a "separate system" which can override linguistic competence at the level of performance. His prediction is that competition effects should be found in differences between comprehension and production, and would be limited to instructed learners, since "naturalistic learners would not have such a separate system." (ibid: 86).

Rothman (2008) compared twenty native speakers (NS), twenty highly advanced instructed learners and eleven highly advanced naturalistic learners on comprehension and production tasks. Overall scores for all three groups found that the L2 learners' scores were very similar to the NS scores, overlapping with the range of individual NS scores (and reaching up to 100% on parts of the comprehension task), and the L2 learners could thus be argued to have native-like competence. However, on elements of each task where there was evidence of variation, statistical analysis found that the instructed learners significantly underperformed the native speakers and naturalistic learners on both tasks in predictable and systematic ways. Rothman argues that this is due to reliance on oversimplified pedagogical rules, which focus on drawing out explicit L1/L2 contrasts rather than drawing on the fundamental grammatical features required. The knowledge of these rules "are consciously accessed in discourse as an output monitor by many L2 learners, resulting in surface morphological errors despite a morphosyntactic competence that is fundamentally native-like" (ibid: 99). In terms of the focus of this study, this hypothesis could be tested using comparison of different types of taught question forms (where asymmetries between simple and embedded questions, or subject/object questions may not be drawn out), and a reliance on a single simplified grammatical rule – e.g. do-support is required in English questions, would produce surface nontarget forms in production (e.g. overgeneralisation of do-support) although the underlying linguistic syntactic features may be in place to generate targetlike judgements in comprehension.

These accounts also connect plausibly with the bimodal models of memory (Schwartz 1993; Ullman 2001), all of which assume that both L1 and L2 systems operate within the multilingual multicompent mind – discussed in more detail below. These accounts argue that L2 is driven by different memory stores or processes to L1, whereas the MOGUL model, I believe, would see the stores and processes themselves as similar but require differing levels of activation.

To bring together the themes discussed in this section on models of mind, I believe that the models discussed so far do not, as yet, provide a full picture of the L2 mind which takes sufficient account of the overlapping nature of L2 knowledge and processing involved in L2 acquisition. I argue that what is needed is a model of mind clearly founded on a coalition of linguistic and processing resources (Herschensohn 1999). I argue that an L2 mind should not be the same as a monolingual mind, but instead should be seen as a "multicompetent mind" (Cook 1991) with overlapping sources of knowledge, in which L2 knowledge is a system "in its own right" (Cook 2002: 9), reasserting the commitment to the notion first raised by Selinker of the independence of the L2 learner's interlanguage system (1972). Calls by Hawkins (2008) for this kind of combined model of SLA can also be found in earlier articulations such as Raupach (1987), Jordens (1991) and Towell et al. (1996) who concluded that "multiple sources of knowledge" may be available for an L2 user; "these sources include setting L2 parameters on the basis of external evidence, transfer of L1 parameter settings, transfer of L1 surface structures (transliterations), the use of [chunks] and of consciously learned rules." (Towell et al. 1996: 90).

Herschensohn (1999) follows this approach in her coalition model for instructed SLA. She argues "that the L2 learner uses a coalition of resources including a UG template, L1 transfer, primary linguistic data and "instructional bootstrapping" (ibid: 220). In this model, she discusses how the language user is able to draw on both implicit and explicit knowledge about language through a "general cognitive awareness of grammatical principles and substantive universals instantiated by L1. This knowledge is presumably located outside the language module in the knowledge base" (ibid: 184-85). She argues (in line with Paradis 2004 and Ullman 2001) that early L2 grammar depends, however, to a large extent on declarative memory (1999: 192), including explicit learning strategies for vocabulary and morphology (Herschensohn 1999: 203).

If this assumption is agreed that declarative or explicit memory is the "scaffolding" (ibid) of the L2 development process, then understanding how explicit and implicit memory is stored and accessed and the role of WM in this process could be the key to understanding L2 development. It could be argued that the reliance on learned or instructional bootstrapping reduces the applicability of this model to naturalistic SLA, but it is also possible that even naturalistic L2 learners develop some kind of monitoring or metalinguistic rule awareness that they utilise (as mentioned by the naturalistic learner investigated in Ioup et al. 1994).

2.9. How implicit/explicit memory fits into development

So far, I have argued that both implicit and explicit knowledge, and implicit and explicit learning, can play a combined role in SLA; I have also argued that noticing or attention may be key to explicit learning, and potentially implicated in implicit learning. However, despite the evidence that implicit knowledge can be acquired, either through invoking UG principles or cognitive inductive mechanisms, there remains divergence in discussion of the distinction between explicit and implicit knowledge, and therefore how these constructs could operate in SLA, and how far it may be valuable to address these questions via "dual-system explanations of human learning" (Robinson 1997: 47).

I now turn to examine the two models referred to at the start of this chapter, underpinning the research questions of my whole study, which overtly address the issue of a two-fold nature of explicit and implicit knowledge in the L2 mind: Schwartz's distinction between learned linguistic knowledge and implicit competence, and Ullman's declarative/procedural (DP) model.

2.9.1. Schwartz (1993)

Schwartz (1993) draws on Krashen's famous distinction between acquisition and learning (Krashen 1985), linking this to the distinction drawn in generative theory between competence and performance. Schwartz suggests that it is important to distinguish two aspects of interlanguage – that which is based on metalinguistic or learned linguistic knowledge (which could be termed "learned linguistic behaviour") and that which is based on implicit competence, seen in naturalistic performance.

From this perspective, she observes that explicit data, such as rule instruction, and negative data do not "fruitfully" impact on L2 acquisition, and argues that they will be "essentially ineffectual in building grammars" (1993: 159). She argues, taking Fodor's (1983) modular model of language, that the language faculty "cannot access the knowledge that gets learned" as a result of explicit and negative data (ibid: 158) – as if there are no correspondence mechanisms to "translate" such learned data into a form the language faculty can process and integrate as competence. Competence is built up, in her view, only by sufficient exposure to primary linguistic data. Schwartz's prime interest seems to be in questioning the "mystery" (ibid) of why different types of input do not have more effect, but by doing so, she highlights the issue at the heart of my research study of considering "what it means to build up knowledge in … long-term memory" (ibid: 159). Schwartz left this question unanswered at the time, and, as we have seen throughout, the answer is still far from clearly understood.

2.9.2. Ullman's Declarative/Procedural Model

Trying to address a very similar question, but starting from an entirely different, non-modular cognitive perspective, Ullman (2001, 2005) has proposed a bimodal model of memory which dovetails with the dichotomies discussed by Schwartz. Ullman extends Pinker's (1999) dual mechanism or words-rules distinction: words are individually learned and stored in declarative memory, but regular rule-generation is based on computational processing. Ullman (2005) applies this distinction to apply to the declarative/procedural dichotomy explained in section 4 above. In his Declarative/Procedural model, Ullman argues that in L1A, the declarative memory system underlies the mental lexicon, while the procedural memory system subserves aspects of the mental grammar (148). Young children initially learn words and forms in declarative memory (149), while the procedural system gradually acquires the grammatical knowledge required for generating rule-based combinations (ibid)[8]. For SLA, his model provides an attempt to explain behavioural and neurological evidence (see also Vos et al 2001) that less proficient learners, whose language use is non-automatic and effortful, rely on explicit knowledge (shown in neural activity patterns associated with declarative memory, e.g. ERPs showing central-posterior negativities, referred to as N400s).

[8]There could be a caveat to this suggestion from other psychological accounts that very young infants' implicit or procedural memory is more developed than their declarative memory (Baddeley et al 2009).

On the other hand, highly proficient bilinguals, who show more automaticity and less effort, show greater similarity to native-language speakers in terms of greater automaticity and similar neural activity patterns associated with procedural memory (e.g. left anterior negativities or LANs). Ullman argues that adult language learners' greater variability in achieving proficiency in L2 reflects maturational changes in declarative and procedural memory systems at the onset of puberty (possibly driven by hormonal changes). While some of these claims remain speculative, his arguments attempt to provide a neurological basis for the evidence argued by proponents of Critical Period effects in SLA, or those who uphold the Fundamental Difference Hypothesis.

Ullman does not discuss the role of WM specifically in his model, but his emphasis on the reliance on declarative, conscious knowledge for less proficient L2 learners, allows the logical inference that WM, the workspace for conscious knowledge, plays a key role in L2 language development and use until sufficient procedural knowledge has been developed.

A bimodal separation of different language knowledge sources has long been argued by Paradis (1994, 2004, 2009), who also addresses the issue of implicitness and explicitness as defined by N. Ellis. Paradis adopts conclusively the learning/acquisition distinction (like Schwartz 1993), providing ample neurological and theoretical justification for his claim that the two systems of knowledge are separate. He stresses (counter to Anderson's 1993 model of increasing automaticity) that explicitly known rules cannot be transformed into implicit computational procedures (2009: 16), and that "nothing outside of implicit linguistic competence can have an influence of any kind on the grammatical system" (2009: 59). He admits that both implicit and explicit knowledge can be accessed automatically (nonconsciously), but that the key difference between the two sources of memory is whether the explicit knowledge can be expressed consciously if required (2004: 43).

Paradis therefore believes that WM can play no part in implicit acquisition, since "all active components of cognitive working memory are declarative, hence conscious ... There is no possibility that inherently implicit and explicit representations can interact in working memory" (2009: 49-50). However, he does recognise that metalinguistic or consciously recorded knowledge can indirectly facilitate implicit acquisition - if explicitly presented sentences are presented frequently, perhaps through practice, this naturally increases the input of primary data to drive the kind of implicit inductive usage-based associations that Paradis sees as underlying acquisition, although he does not, as far as I understand it, go on to define what kind of input is needed to drive successful acquisition.

All the bimodal models discussed here rely on a distinction between two types of knowledge, and yet issues remain over how such a distinction can truly be identified, measured and tested, despite the many studies that have tried to do so (only some of which have been discussed here). There are also some neurological difficulties with the assumption that the two systems of declarative knowledge and procedural competence are used in different ways at different ages and stages of proficiency. Counter to the processing differences found by Clahsen and Felser and others (see section 2.7.4.3), and to findings referred to by Ullman (see above), other studies found evidence of similarity in processing. Perani et al (1998) found evidence that highly proficient learners showed similarity of brain activation to native language users, regardless of age of acquisition. Green (2003, 2008) puts forward the convergence hypothesis that growing proficiency would indeed lead to L1 and L2 knowledge being processed the same way, but suggests that "we have little or no information about the functional integration of different neural regions during second language use" (2003: 212). Moreover, there are questions over the wider implications of this kind of neurolinguistic methodology. De Bot (2008) warns against too much reliance on neurological tests which often only look at specific words or short phrases, as providing evidence to justify theories of language learning more generally.

2.10. Justification for WM's role in L2 grammatical development

It is clear that many issues remain outstanding, including how to define LTM in SLA, and how WM interfaces with LTM, given an assumed dichotomy within LTM of the distinction between implicit or procedural knowledge, and explicit or declarative knowledge. If WM is the workspace for knowledge that is controlled and can be consciously held, it is easy to infer that WM is primarily concerned with explicit, declarative knowledge. So whether WM plays a role in SLA depends on how far SLA is driven by implicit or explicit learning. Explicit learning, for this purposes of this study, is associated with instructed presentation of grammatical chunks, negative feedback and conscious awareness of rules about language, commonly found in traditional grammar-instruction classrooms, and role-drilled learning, as commonly found in Chinese classrooms (Gu 2003). Implicit learning, where conscious awareness is not paid to the input, is associated with child language acquisition and naturalistic second language learners (especially low literate learners – Tarone and Bigelow 2005; Craats et al 2006). It is argued that WM, for these kinds of language learners, does not play a major role in implicit morphosyntactic acquisition (Juffs and Rodriguez 2006). This is most famously expressed in Newport's (1990) "less is more" hypothesis, where younger children's lower levels of working memory entail greater implicit learning, compared to post-puberty children and adults' greater working memory capacities, which could override implicit mechanisms, in favour of explicit problem solving and use of longer memorised chunks, that may well have been the major source of language input for many instructed learners.

However, the wide range of studies both within the generative and cognitive frameworks evaluated above suggest some element of memory and cognition is involved in language development, since maturing processing capacity, and, arguably, growing WM capacity, is seen by both cognitive researchers (Tomasello 2003) and generative researchers (Rizzi 2005) to be crucial for acquisition by children of complex syntax, e.g. CP phenomena such as wh-questions. In addition, given that complex linguistic tasks found in many aspects of language processing, production and comprehension involve marshalling linguistic knowledge in real time (Jackendoff 2002: 206-230), WM is logically central to language use as well as acquisition.

As I have argued throughout, L2 interlanguage for instructed learners reflects a mixture of explicit knowledge - both of words, chunks, and knowledge of grammatical rules and, arguably, much inflectional morphology (Emonds 2000) - and of nontarget-like implicit knowledge (either in the form of very underspecified L2 features or stored with insufficient levels of activation to overcome L1 levels of activation). I have discussed a number of models which assume at least two sources of linguistic knowledge in some sort of combination. However, one of the chief difficulties with the bimodal models discussed above is that they do not effectively provide an account of how transition is achieved in changing reliance on one system to reliance on another. I assume that for learners who cannot yet rely on well established (automatic) competence, and still show extensive variability, the need to comprehend input or produce output leads to many occasions where there is failure to code automatically - at this point, the different sources of knowledge must operate in competition with each other for priority (Felix 1985; Truscott and Sharwood Smith 2004; Truscott 2006; Rothman 2008).

Given the ensuing competition between L1 specifications, UG constraints, underspecified L2 nodes and items in the lexicon (available both as declarative and procedural information), conscious control and inhibition of unwanted information will be needed in processing input and output, which are two of the central roles argued for WM. In addition, if acquisition is failure-driven (Carroll 2001), then WM is potentially the means for hypothesis-generation for possible parses of input or creation of output where failure requires a solution.

Research trends in WM research (utilised by Truscott and Sharwood Smith, for example, and also acknowledged by Baddeley 1999; 2007) is that working memory should not simply be seen as a workspace used for conscious manipulation of consciously learned declarative knowledge, but a set of activation and retrieval processes, in which all of LTM can be activated (i.e. declarative AND procedural information). However, I assume that WM will prioritise retrieval of declarative knowledge since this seems to accord most closely to standard psychological models of LTM and STM storage and retrieval (Smith and Kosslyn 2007).

In terms of WM in SLA, investigating acquisition and use of different constructions (e.g. simple questions that have been explicitly taught, and are frequent in the input, versus complex questions and subjacency violations which are less frequent or impossible in the input) would provide empirical evidence of how differences in WM capacity would affect individual differences in acquisition, and potential asymmetries between taught and non-taught constructions. There are two key perspectives to investigate this, focusing on WM in language use and WM in language development.

2.10.1. WM as workspace for existing knowledge

The first question is whether greater WM capacity facilitates better use of existing declarative knowledge. Two types of tasks could be used to test the hypothesis that L2 acquisition and processing uses declarative knowledge, mediated through WM: oral output and grammaticality judgements. Firstly, in oral output, declarative knowledge processing could be accurate but slow, or effortfully produced, containing slips and restarts as a result of monitoring form as well as content; structures would be produced with variable success (Dechert 1980; de Bot 1992; Towell et al 1996; Temple 1997; Kormos 2006). This would be because as speech is formulated, products that would normally be stored in the syntactic buffer or articulatory buffer, which are beyond attentional control, are instead routed to the limited working memory store (Kormos 2006). As a result, formulation proceeds slowly and serially rather than quickly and in parallel fashion. This shows up overtly in slower speech rate and in more hesitations, particularly mid-clause hesitations (Temple 1997: 86-87). Greater WM capacity should facilitate faster formulation and retrieval, revealed in faster, more accurate speech.

Secondly, grammaticality judgements could reveal asymmetries in accuracy between structures that would have been explicitly taught (indicating the degree of reliance on declarative memory), compared with ungrammatical non-taught complex forms, such as subjacency violations. Grammatical forms available in the input should be judged more easily than ungrammatical forms, which must be analysed whether they are grammatical but unfamiliar, or ungrammatical and therefore impossible. Greater WM capacity again should facilitate more accurate judgements and how easily judgement tasks are carried out.

2.10.2. WM as key to faster development of new knowledge

There is a second question about the role for WM which is relevant in studying language development in the context of changes in type of exposure. By testing the changes in accuracy and fluency as the amount of exposure increases, for example in an immersion setting, it should be possible to assess how far immersion aids faster development and greater accuracy of explicit knowledge (Howard 2006), and whether WM facilitates this by providing the capacity to respond to forms which may be increasingly frequent in an immersion context compared to an instructed context, and marshall the output more efficiently. It has been shown above in section 2.7.4.4 that there are contrasting findings from studies on language development in immersion or study abroad settings (Freed 1995; Sunderman and Kroll 2009), but existing studies are not clear as to why that should be so. I argue, following Schmidt's approach, that working memory is required for providing conscious attention or "noticing" the structure in the input which, if sufficiently enriched or rehearsed, would facilitate acquisition of that structure (Ullman 2005; Paradis 2009). Therefore working memory capacity is argued here to play a part in L2 development in the context of enhanced exposure as provided by immersion.

A further question arises from the effect of greater exposure on the asymmetries proposed above. If WM is associated only with conscious noticing and with facilitating explicit knowledge, as hypothesised above, then comparing longitudinal development on explicit taught common forms with development of sensitivity to implicit untaught subjacency violations should provide evidence of how WM facilitates different types of knowledge develop. Evidence of significant differences in correlations between working memory and higher scores or faster times on taught versus untaught structures would confirm that working memory facilitates explicit learning.

2.11. Empirical studies of working memory in L2

Despite the logical assumption that WM is necessary to L2 development, particularly within the cognitive paradigm of SLA (Robinson 2002), WM research in L2 has found differing results, suggesting either that the assumption is incorrect, or that that current WM tests are not valid reliable tests for L2, or that they do not necessarily interact with linguistic data as they are assumed to do. It is therefore critical to establish what kind of L2 data would be best for analysis, and how different WM tests are thought to affect this data.

Most of the studies on WM in L2 have focused either on tests of phonological loop capacity to predict vocabulary learning or oral fluency (Baddeley et al 1988). Tests usually focus on non-word repetition (Gathercole and Baddeley 1993, Gathercole 2006), which measured accuracy in repeating novel sounds of non-words of increasing length (two syllables, three syllables and so on).

Other studies have been prompted by the research into executive control (by Daneman and Carpenter and others in L1) predicting greater skill in reading and listening comprehension from scores on the Reading Span Test (Daneman and Carpenter 1980) and the version developed for Listening Span (Daneman and Green (1986), which were outlined in section 4.5. To recap briefly here, the Span test is seen as a test of the storage-processing trade-off assumed to operate in a fixed capacity system, where greater amounts of information to be stored offset the efficiency of processing – greater executive capacity means that greater storage (or faster processing) is found. The Span test (in either form) consists of a certain number of sentences being repeated or judged for truth value in blocks of increasing lengths (e.g. five blocks of three sentences in each block, then five blocks of four sentences, five blocks of five and so on); at the end of each block, participants have to recall accurately and in order the final words of each sentence. Once a participant cannot recall the final words on two or more blocks out of all the blocks presented in that set, the test usually ceases.

Tests of both simple phonological storage and executive storage/capacity trade-off were used in a seminal study of WM in L2 by Harrington and Sawyer (1992). Using a modified L2 version of the Daneman and Carpenter Reading Span test, they investigated advanced Japanese learners of English, and found significant correlations for WM with L2 reading ability and grammar performance on scores in the reading and grammar sections of the TOEFL standardised proficiency test Simple storage measures of English digits and words did not show any correlation on either the grammar or reading test scores, suggesting that it is the involvement of the central executive element that is crucial in L2 processing. This may be explained that the two tasks within the TOEFL both measured explicit knowledge, since reading is an explicitly learned skill, and the type of grammatical proficiency measured in TOEFL tests can typically refer to learned grammatical knowledge, or can be successfully answered using metalinguistic knowledge of explicitly learned rules.

Osaka and Osaka (1992) also investigated Reading Span in both L1 and L2 for adult Japanese learners of English, and found that WM capacity scores as measured in L1 and L2 correlated highly with each other (r=.72). This finding led to the robust assumption that WM span is not language-specific.

Ando (1992) followed a group of younger Japanese learners of English (6th grade) through 20 hours of instruction from beginner level, in a traditional grammar-oriented approach. The children's Reading and Listening Spans in their L1 measured before the teaching instruction began were the strongest predictors of their posttest performance after the period of instruction ended ($r=.60$, $p<.05$ for Reading Span, and $r=.72$ for Listening Span) among a range of cognitive and personality measures administered.

Two other studies specifically addressed WM and grammatical acquisition using Reading Span tests. Miyake and Friedman (1998) tested 59 adult Japanese learners of English for their semantic and syntactic comprehension of thematic role, using the Competition model of cue-preference (referred to in section 4.3 above). They tested how far Japanese L1 use of morphological particle and animacy cues could adapt to English L2 use of word order to cue subject role. Miyake and Friedman argue that more global cues like word order demand higher WM than local particle-type cues. WM capacity was tested in English and Japanese listening span tasks, and Miyake and Friedman found that both L1 and L2 WM scores correlated significantly with accurate syntactic comprehension ($r= .49$ for L1; $r= .52$ for L2, $p< .01$ for both). They analyse the multicomponent model of WM to infer that the combination of phonological loop storage and central executive processing provide the mental workspace for the kind of analytic ability and memory capacity that are theoretically assumed to make up language aptitude (Skehan 1998), and propose that "WM-as-language-aptitude" may well be the key to variation in L2 learning and use (1998: 361).

Robinson (2002) also investigated WM using Reading Span to look at the role of WM in different types of learning context, to see if WM was implicated in incidental learning, which he equates to Krashen and Schwartz's definition of acquisition. Robinson argues that WM does play a role in L2 but in a task-specific way, so that correlations would be conditioned by task complexity and processing mode involved in the test of any given language structure. He tested incidental learning (of novel Samoan vocabulary and different grammatical rules for ergative or locative structures), by 38 adult Japanese learners, over a period of a week's treatment. L2 learning was measured using untimed written grammaticality judgement tasks and a sentence production test, immediately after the treatment, followed by a posttest one week later, then a delayed posttest six months later (in which only 26 participated). Robinson ran a range of tests for individual differences, including WM in L1 using Osaka and Osaka's Japanese Reading Span test (but L2 WM was not tested). Robinson found no overall consistent significant positive correlation between WM and accuracy on the written grammaticality task or the production test at any of the times of testing, and neither did his other ID tests.

He initially suggests that incidental learning of grammatical rules "appears unaffected by individual differences" (2002: 259). However, he qualifies this by highlighting individual instances of significant correlations between WM and one specific ergative construction ($p<.05$), as well as nonsignificant but weakly positive correlations between WM and two locative structures which were evident in a timed aural grammaticality judgement task measured in the immediate delayed posttest (one week after treatment). He concludes that "adult incidental L2 learning is variably affected by the complexity and nature of the rule to be learned" (2002: 260), and that individual differences will constrain even incidental learning as long as they are "relevant to the processing demands of that particular learning task or condition" (2002: 262).

In addition to the studies looking at WM in grammaticality judgement tasks, a number of studies have tested the correlation between WM and L2 in oral fluency tasks (Temple 1997; Fortkamp 1999). Temple (1997) argues that WM limitations in L2 hamper the capacity for parallel processing required in fluent speech, resulting in slower, word-by-word serial processing and production. Fortkamp (1999) tested sixteen advanced L2 English learners (mean age 27.5) at a Brazilian university. Fortkamp measured Speaking Span, as tested using Daneman's (1991) speaking span test where sentences have to be produced from a set of recalled words. The span scores were then compared against scores of L2 fluency measured in an oral generation task, describing a picture of a domestic family scene for 1 minute, 30 seconds. In order to test how far WM may be task specific, Fortkamp also measured Reading Span and oral reading proficiency (reading aloud a short 320-word literature extract as fast as possible), and again found significant correlations between these two scores. Fortkamp concluded from these data that WM is task specific (Turner and Engle 1989).

Other studies have focused on the role of the phonological loop, which has traditionally been measured using non-word repetition (Gathercole and Baddeley (1993) and is robustly supported by evidence from L1 studies implicating greater loop capacity with vocabulary learning in school age children (as discussed in section 5.5 above). Evidence suggests that phonological loop capacity significantly correlates with foreign vocabulary learning in adults (Baddeley et al 1988; Baddeley et al Papagno 1998). Service (1992), Service and Kohonen (1995), Cheung (1996), Masoura and Gathercole (1999) also found for primary school children that higher scores in non-word repetition predicted better results as measured in school tests of L2 English writing and vocabulary.

Ellis and Sinclair (1996) found that the simple act of rehearsing specific grammatical structures can facilitate faster learning. In their experiment, for Welsh soft mutation (where certain phonemes change in certain grammatical contexts), a group of participants who repeated aloud the tokens used in the test were significantly better than other groups ($p<.01$) in judging and producing examples than a group of participants who remained silent and a group whose articulatory loop was suppressed.

Kormos and Safar (2008) investigated the effect of differences in learner level or WM test type on a range of proficiency measures. They tested 100 beginner and 20 pre-intermediate adolescent learners of English looking at their proficiency after a year's instruction measured by the Cambridge First Certificate exam, which included reading, listening, composition and an oral Use of English task. Kormos and Safar used an L1 non-word repetition task which only had significant correlation with language test scores for the pre-intermediate learners of English ($r= .47$, $p<.05$ for overall proficiency score, $r= 0.49$, $p<.05$ for oral task). A subgroup of 45 of the original group of beginners were also tested after a second year of instruction using backward digit span (a measure of overall WM capacity, not just phonological loop), and correlations between this task and proficiency scores task was highly significant ($r= .55$, $p<.01$). Kormos and Safar comment on the difficulty of assessing precisely what type of L2 knowledge leads to higher scores on this kind of language exam, since there is no separation between accuracy, complexity, and, in the oral task, global fluency. Similarly to Ellis and Sinclair's (1996) study, they infer that explicit presentation of words and rules support learning of chunks and sequences – higher phonological loop capacity makes it easier to acquire a wider vocabulary, and higher general WM capacity makes it easier to memorise chunks promoting inductive analysis of new grammatical rules.

O'Brien et al (2006), like Fortkamp (1999), looked at WM and gains in oral L2 proficiency, but O'Brien et al used a serial non-word recognition task of phonological memory (adapted from Gathercole et al 2001). Their participants were followed during a study abroad period of three months, and L2 improvement was measured in Oral Proficiency Interviews including vocabulary, narrative abilities, accurate use of grammatical morphemes and use of subordination (detailed in Collentine 2004). Their results were mixed, in that higher phonological memory scores were not implicated in greater vocabulary use, but were significantly associated with better narrative abilities, and greater grammatical accuracy.

Sunderman and Kroll (2009) also focused on the interaction between WM and study abroad, looking at developments in lexical proficiency, tested through word comprehension and naming tasks. Forty-eight L2 learners of Spanish were tested, 34 of whom had studied abroad (SAE), but all of whom had had similar lengths of school study, and were at similar levels of proficiency according to a self-rated scale of 1-10. The comprehension task consisted of a computerised timed word translation task, in which participants had to judge L1 and L2 pairs for accuracy of translation as quickly as possible by pressing a button for yes or no. The picture naming task again was based on a computerised timed task, with a new picture generated by pressing a button after the picture had been named.

Linguistic scores were correlated with a Reading Span task for working memory span based on Caplan and Waters (1996), in which participants judge sentences for semantic plausibility, which are blocked into sets of increasing size (two to six sentences long), and final words of each block are recalled after each block is presented. Sunderman and Kroll found a clear difference between the mean scores of the SAE group and the other group in the translation task and the production task, showing an effect of study abroad on accuracy and processing speeds. They also found that those with higher WM were also more accurate and faster in processing speeds, regardless of SAE, but only for the translation task. The interaction between WM and SAE was less clear; in a highly complex statistical analysis, it was found that only those with higher WM showed any significant effect for WM on improvement on both tasks. Sunderman and Kroll conclude that there might be a kind of internal threshold of WM capacity, above which study abroad does facilitate improvement, especially in comprehension, but below which added exposure through study abroad does not make a significant difference.

Mackey et al (2002) also found mixed evidence for the role of WM in linguistic development, using non-word recall and listening span. Mackey et al (2002) followed up an earlier study investigating the role of feedback and interaction (Mackey 1999, referred to in section 5.15), and specifically addressed the question of whether WM facilitates greater effect on noticing feedback and improvements in question formation. The (2002) study investigated a group of 30 Japanese study abroad learners, of intermediate standard (measured in TOEFL scores) and with a mean length of residence in the US of 8.6 months, who were given interaction-based treatment sessions lasting a week, with a delayed posttest two weeks later. WM was measured using nonword recall (Gathercole et al 1999) and Listening Span tests in English and Japanese (Osaka and Osaka 1992; Waters and Caplan 1996). Noticing was measured using stimulated recall. Mackey et al found that participants with lower WM noticed feedback less than those with higher WM, but only in terms of the nonword task. There was no correlation between WM and noticing, as measured by stimulated recall, but there was some evidence in a very limited subset of the group (only six) that higher WM seemed implicated in linguistic development through the stages of question formation. There were only thirteen participants in all involved in the final stages of analysis, and Mackey et al are careful not to make over-generalised claims, only stating that their data only indicate a "potential relationship" (2002: 204) between WM and L2 development.

By contrast to the studies discussed above, a number of studies have not found any effects for WM across a range of L2 issues, including processing, grammatical development, or oral fluency; three studies are reported on here which had an impact on the hypotheses and study design of my research study.

In a study looking specifically at parsing in L2 for higher-level learners, Juffs (2004) investigated the evidence found in L1 research (Miyake et al 1994) that WM based on Reading Span correlated with faster thematic assignment when processing garden path subject/object ambiguity structures such as "Before the student guessed the answer appeared on the next page". Juffs tested 104 adult upper-intermediate participants (with Chinese, Japanese and Spanish L1s), assessing online processing load at possible interpretation sites through a moving window technique by measuring speed and accuracy; participants' Reading Span scores were measured in L1 or L2 and a storage-only word recall task was also done in L1 and L2. However, unlike Miyake et al (1994), Juffs found no correlation between any of the WM measures and either accuracy or speed on the linguistic task.

In a semi-longitudinal study of WM and L2 development, Sagarra (2000) tested 110 early and late beginner learners of Spanish looking at grammatical development over a period of 18 months' instruction in a wide range of introductory Spanish morphosyntax and function words (such verb morphology, auxiliaries, adverbs), measured by university grammar tests, including written sentence completion tasks, cloze tasks and grammaticality judgement tests. WM was measured using the Daneman and Carpenter (1980) Reading Span. She found no correlation between linguistic scores and WM as measured in Reading Span.

Towell and Dewaele (2005) conducted a small-scale longitudinal study of twelve advanced learners of French measured prior to and after a period of study-abroad immersion in the L2 environment, which was particularly interesting given the scope and aims of my research study. Towell and Dewaele focused on the interaction between competence, automatisation and oral fluency. Grammatical awareness (competence) of clitic pronouns, use of negation and the use of adverbs; oral fluency was tested in retelling a short cartoon story. Oral fluency in the context of working memory was tested using a "shadowing task" involving immediate and continuous repetition of what was being said (i.e. repeating one part while listening to the next). Oral fluency in the context of automatised knowledge was tested using a four-sentence mini-story repetition task, designed to be too long to be held in the phonological short-storage element of working memory. However, they found no correlation between WM, measured in a shadowing task, and developments in grammatical accuracy or oral fluency

Mizera (2006) investigated the claimed connection between L2 oral fluency and WM, using Reading Span, but found, counter to Fortkamp (1999), no correlation between

Reading Span and fluency or complexity in L2 speech in a group of 44 adult learners of Spanish of lower and upper intermediate level learners.

It seems clear that the evidence remains mixed in investigations of the role of WM in SLA, including longitudinal development of oral fluency or morphosyntax, which is of particular interest for my study. One issue arising from this contradictory evidence is that there is still some controversy over how to assess the role of WM in L2 (Sagarra 2000). Some of these difficulties may arise from the traditional word storage and recall measure of Daneman and Carpenter's Reading Span Test, which may simply be a circular issue of a reading test measuring reading proficiency (Sagarra 2000; Mizera 2006). In addition, in view of the time lag involved during the test procedure between storage and recall, especially on the longer sentence sets, word recall may in fact be using some other kind of short-term memory rather than the phonological loop (Mizera 2006: 17). The phonological loop is assumed to be around 1-2 seconds which would not be long enough to hold the information required in the longer sets of sentences. It is possible that successful recall could be tapping into some element of long-term working memory (Ericsson and Kintsch 1994), or it might be evidence of Baddeley's suggested episodic buffer (Baddeley 2000), although this is purely speculation.

Very little research on the proposed episodic buffer in language has been done, and none in the context of L2 learning. Fehringer and Fry's (2007) small-scale study on bilingual use of complexity in spontaneous speech (see section 2.6.5) is, as far as I am aware, the only study which has to date attempted to tap this construct, using a psychologically established Story Recall task, although Towell and Dewaele (2005) reported above used a somewhat similar mini-story repetition task.

The interaction between task demands and the different elements of the WM construct also need further investigating, and it seems clear that Daneman and Carpenter's Reading Span Task (or the Listening Span variation) may not be the best tool for WM research, particularly in L2A. There is also considerable discrepancy across the actual ways of measuring WM, which is notoriously unreliable, even in L1 (Waters and Caplan 1996; Conway et al 2005), which also indicates that care must be taken in how WM should be tested. In addition, although WM in adult native speakers is assumed to be stable (Gathercole and Baddeley 1993; Meyer et al 2001), it is also noted that measures of memory, including WM, can vary significantly depending on affective external or internal factors, such as time of day, hunger, tiredness, changes in hormonal levels (Matthews et al 2000). None of the studies of WM and L2 I have examined appear to have taken this into account.

In conclusion, after this review of studies of WM in L2, a number of gaps and mixed findings have been established in the current extent of WM research in L2, which this study goes some way to address. There does appear to be a consensus that WM can affect both language use and development, despite methodological issues in identifying suitable linguistic and WM measures. Nonetheless, the role of WM in syntactic acquisition and variability in morphological production is unclear, compared to lexical learning or explicit rule retrieval, and no study has yet sought to investigate the full range of wh-constructions, which have been identified in this study as liable to cause difficulty in L2A.

Towell and Dewaele (2005), and Sunderman and Kroll (2009) also provide interesting mixed findings on how language changes as a result of study abroad/ immersion, and the interaction of immersion with WM, which was of particular impact on my study.

My study thus aims to add conceptually and empirically to the debate over how far WM can be shown to impact on language acquisition and use. By using a battery of WM tests, including tests of the episodic buffer, I hope to increase our understanding of how WM operates in the L2 acquisition of syntax, and avoid issues of unreliability and lack of task-appropriateness.

2.12. Motivation for this study, research questions and hypotheses

To conclude, this chapter has argued that UG plays a role in developing L2 linguistic knowledge, but other sources of knowledge and processing constraints also affect L2 linguistic development, especially in explaining inter-learner and intra-learner variability. Therefore modular automatic unconscious processing, using UG-based combinatorial rules, is possible, but difficult and rarely fully achieved. It is further assumed that the starting point for most instructed L2 development is conscious learned knowledge of that L2, in conjunction with implicit procedural knowledge and explicit metalinguistic awareness of the L1. It has been shown that even when implicit linguistic competence appears to be in place, processing issues affect online management of the complex task of sentence comprehension and production. Even at advanced levels of L2 knowledge, online linguistic processing requires the kind of controlled attention and manipulation of knowledge resources that requires WM. Thus I argue WM capacity can affect L2 development throughout acquisition, and may even be the key to variation in L2 development (Miyake and Friedman 1998).

One way to investigate the role of WM is to test the assumption that individual differences in working memory constrain the capacity to store novel information and retrieve existing knowledge, particularly explicit knowledge (Baddeley 2003). I want to examine the assumption that Chinese speakers of English, even at advanced level, rely primarily on explicit (declarative) learned knowledge, evident through slower, more monitored, more hesitant, more variable processing/production (Segalowitz 2003; Ullman 2005; Paradis 2009). If this is so, the impact of differences in WM capacity is predicted to be reflected in individual differences in learners' use of their existing explicit knowledge, which has been shaped by an input environment which is poor on naturalistic native input or interaction, and has prioritised memorisation of learned sentence exemplars and explicit grammar rules (Gu 2003).

It has also been established that changes in input environment may play a key role in affecting acquisition but that individuals may vary in their response to input. It is assumed that immersion will provide added exposure, triggering L2 development of greater automatic accuracy and less variability or effort (Howard 2006), but it needs to be clarified whether this is found more for explicit taught constructions than for implicit untaught constructions. A further aim of this study is to investigate how far linguistic development in these target structures would be affected by WM capacity in response to the new richer input environment, testing the assertion that WM is the key to variation in L2 (Miyake and Friedman 1998).

Question forms have been identified as an appropriate group of constructions to test these questions and assumptions. To recap the issues discussed in earlier sections, fully inflected questions, especially wh-questions, are argued to be late acquired (Pienemann 1998) due to higher processability constraints, and may well also reflect certain inherent syntactic difficulties at CP level (following the suggestions of Platzack's 2001

Vulnerability C-domain hypothesis, raised in section 2.2). The features of wh-movement and head movement in English are predicted to cause difficulties for speakers of wh-in situ languages such as Chinese, in acquiring the relevant features to trigger wh-movement, as well as asymmetries between subject and object questions, and simple and embedded questions. In addition, constraints on long-distance movement which are argued to apply differently between Chinese and English are predicted to cause difficulties either from processing limitations or through lack of access to the relevant formal principles of subjacency.

Given the underdetermination in the input for some of these constructions (especially subjacency-constrained structures), research into their acquisition by learners whose L1 lacks wh-movement has been seen as critical to the issue of whether UG was accessible for L2 acquisition (White 1989). Simple and embedded questions are widely taught in a formal setting (at beginner level, e.g. in Taiwanese junior high school classes, and again at intermediate level, such as Cambridge First Certificate/IELTS 4), but the difference between subject and object questions is not usually explicitly highlighted in the way questions are presented (Acklam 1996; Nani 2006). The constraints involved in subjacency are usually taken to be "unlearnable" in explicit terms (White 2003: 22), since ungrammatical expressions are not produced in the input. The issues in acquisition of wh-movement for Chinese L2 learners of English therefore focus on how successfully they can acquire language-specific features for overt wh-operator and head movement in short-distance questions, for embedded questions, and constraints affecting long-distance movement.

I therefore argue that Chinese learners' knowledge of question formation is appropriate for analysing these research issues for the following reasons.

1. Question formation, for these learners, uses both taught or explicit knowledge (short-distance and indirect questions) and implicit knowledge (subjacency constraints). Question structures are asymmetric in processability between short-distance and long-distance questions, and object and subject questions. Tests of oral production and grammaticality judgements would provide information on acquisition of these structures, showing effects of knowledge

source, effect of mode of processing, and effects of asymmetries in structure type.

2. Question formation provides a crosslinguistic contrast through parametrical differences between wh-movement in English, vs. wh-in situ in Chinese, and existing research into Chinese learners of English has found contradictory findings in the level of their proficiency in question formation and knowledge of UG constraints.
3. Chinese learners' exposure to English is argued to be impoverished, with minimal practice in online automatic L2 processing in the traditional grammar translation approach to English teaching in China (Gu 2003). The effect of immersion should therefore be more apparent.

This investigation is thus intended to contribute both to the theoretical and empirical questions posed by WM test design and to a greater understanding of how WM may affect the processes involved in language development and variation as outlined in the following four hypotheses:

Hypothesis 1. Instructed Chinese learners of English will show asymmetries in their knowledge and use of different questions forms, in that taught (explicit) simple questions will be more targetlike than either complex questions or untaught (implicit) subjacency constraints, and that object questions will be more targetlike than subject questions, measured in oral output and written tasks.

Hypothesis 2. These learners will improve in their knowledge and use of question forms (subject to the asymmetries noted in 1) when they are exposed to enriched input in an immersion setting.

Hypothesis 3. WM capacity is implicated in individual differences between learners' abilities to access existing knowledge of question forms efficiently and in differences in rates of development from simple to complex questions, measured in oral output and written tasks.

Hypothesis 4. WM capacity is not implicated in their capacity to acquire untaught implicit subjacency constraints, measured in written tasks.

Chapter 3: First Study - methodology and findings

3.1. Introduction

This research was designed as a longitudinal empirical investigation into the correlation between Working Memory (WM) and L2 variation in the acquisition and production of question forms shown by adult Chinese speakers of English in an immersion setting during a study abroad period. This design arose from evidence of morphosyntactic asymmetries in the development of English question forms for advanced learners, and the potential role for WM constraints as a factor explaining individual differences in these asymmetries, as discussed in Chapter 2 (e.g. sections 2.2, 2.6, 2.8, 2.10). The research questions motivating this investigation led to four hypotheses, posited at the end of my literature review (2.2.12), and tested in the first study discussed in this chapter. The four hypotheses are repeated here for ease of reference:

Hypothesis 1. Instructed Chinese learners of English will show asymmetries in their knowledge and use of different questions forms, in that taught (explicit) question forms will be more targetlike than untaught (implicit) subjacency constraints, and that object questions will be more targetlike than subject questions, measured in oral output and written tasks.

Hypothesis 2. These learners will improve over time in their knowledge and use of question forms (subject to the asymmetries noted in 1) when they are exposed to enriched input providing primary linguistic data in an immersion setting.

Hypothesis 3. WM capacity is implicated in individual differences between learners' abilities to access existing knowledge of question forms efficiently and in differences in rates of development from simple to complex questions measured in oral and written tasks.

Hypothesis 4. WM capacity is not implicated in their capacity to acquire untaught implicit subjacency constraints, measured in written tasks.

It was important to ensure that WM was the principal independent variable under investigation, given that so many factors play a role in L2 variation, particularly L1 transfer, age and type of exposure (Flege and Liu 2001; Moyer 2004). Care was taken to ensure that these potential confounding factors were as homogenous as possible in the study design described here.

Since the combination of formal morphosyntactic and WM research questions investigated here was unusual, it was critical to ensure that the pioneering combination of tasks and methodology in this study was valid. A first study was thus carried out as a full-scale longitudinal pilot, which is described in the rest of this chapter. Section 3.2 gives details on the linguistic tasks and data collection procedure, and section 3 covers the WM tasks. Section 3.4 details the participants; section 3.5 details the data collection procedures, including methodological problems that arose. Section 3.6 provides results and discussion to explain how the tasks, methods and hypotheses were revised for a more extensive second study. Tasks are detailed in full in Appendix A. The second study is presented in chapters 4 and 5.

3.2. Tasks used in the first study

In order to test the hypotheses outlined above, a battery of linguistic and WM tasks was constructed, shown in tables in section 3.2.1 and section 3.3 below. These tasks were administered to participants three times, firstly within two months of arriving in the UK (see section 3.4), then again after six months' immersion and finally after around ten months' immersion (Time 1, Time 2 and Time 3). The data collection procedure was designed to be administered in a group format in order to maximize efficient data collection for statistically adequate sample sizes (see section 3.5). As discussed in the literature review (section 2.7.1), cross-modal data collection procedures were used to ensure evidence of acquisition could be triangulated between oral and written data to test for evidence of developmental stages and/or syntax-morphology impairment.

3.2.1. Linguistic tasks

To test hypotheses 1 and 2 about asymmetries in L2 English question formation, three tests were used to measure question production and grammaticality judgements (an oral question and answer 2-way gap fill task (Task 1), adapted from Mackey (1999), and paper grammaticality and production tasks (Tasks 2 and 3, adapted from White and Juffs 1998).

3.2.1.1. Task 1: Oral task

Task 1 was designed to provide a global measure of use of simple and complex question forms, by measuring the number of targetlike question forms spoken over a 7-minute "spot-the-difference" exchange between pairs of Chinese learners of English who had to complete and match their semi-filled pictures of a party scene (adapting a similar task used by Mackey 1999) – see Appendix A.

Table 2: Overview of linguistic tasks

Linguistic Tasks	Focus	Procedure	Measure
Oral question task: Question total Question ratio	Semi-spontaneous question forms, following Pienemann's stages of development	Digitally recorded paired gap-fill activity using pairs of semi-completed pictures to "find the difference", lasting around 7 minutes.	Question total = total number of targetlike questions at higher stages of development. Question ratio = total divided by total number of utterances (between 0 and 1).
Written grammaticality judgement task	Time-limited judgements of object and subject long-distance movement and subjacency violations	Each participant instructed to complete as quickly as possible but within 10 minute limit.	Total number of targetlike judgements of grammaticality (out of 22).
Written question production task	Time-limited generation of questions to test knowledge of long-distance movement and subjacency constraints	Each participant instructed to complete as quickly as possible but within 10 minute limit.	Total number of targetlike questions produced (out of 15).

In line with Pienemann's (1998) hierarchy of acquisition, question forms produced by the participants were divided into two groups: Stages 1-3 (formulaic chunks, intonation only, question word fronting without head movement; double marking of verbs), and stages 4-6 (copula fronting and inversion after wh-questions, head movement and "do"-support, cancelling inversion in embedded clauses). Questions from stages 4 and 5 are seen as simple forms; those from stage 6 are seen as complex forms. All utterances were tagged as question or not, and questions were coded according to the stages of development shown in the table below (repeated from section 2.7.1, in chapter 2).

Table 3: Stages of development in L2 English question formation

Stage	Formation	Example
1	Rising intonation on words/ formulae	Four children?
2	Rising intonation on clauses	The boys throw the shoes?
3	Placing question word at front of clause without head movement; double verb marking	Is the picture has two planets on top? Where the little children are?
4	Copula fronting and inversion after wh-questions	Is there fish in the water? Where is the sun?
5	Head movement of auxiliaries, modals, "do"-support	Can you tell me? What is the boy eating?
6	Non-movement in embedded questions	Can you tell me what the date is today?

The key global measure, question total, was the total number of targetlike question forms reflecting stages 4-6 in the table shown above, which I took to imply acquisition of wh-movement. This was in order to test for evidence of acquisition of English feature strength, outlined by Adger (2003), and Radford (2004) as generating target-like accuracy on wh-movement, and asymmetry of head movement in matrix or embedded questions (as discussed in section 2.2 in chapter 2). Targetlikeness was defined as how many questions showed accurate verbal morphosyntax and correct fronting of the wh-word. Nontarget-likeness was defined as errors with verbal morphological inflections of tense or person – evidence of lack of wh-movement or lack of head movement were coded as tokens from stages 1-3. The overall question total was also subdivided by question stage, to get more detailed information about use of questions at each stage of development, and how much progress would be found from lower to higher stages over the year's immersion, to provide data on evidence of development from simple to complex question forms in line with hypothesis 2.

A second global measure, question ratio, was also calcuated to check for the degree of task avoidance (Schachter 1974). Question ratio was calculated by dividing the question total described above by the total number of utterances, to see how far participants' output reflected their capacity to address the task, without reliance on statements or other circumlocution. The total number of utterances was counted (not including non-propositional back channelling such as "ok", "sorry", or filled pauses such as "mmm"), and the proportion of question to utterance was thus calculated as a ratio (measured between 0 and 1).

3.2.1.2. Tasks 2 and 3: Written tasks

It has been noted that oral data may be difficult to analyse if issues of variability in oral production obscure a clear assessment of underlying competence, by entangling linguistic knowledge from what may arise from non-linguistic issues or "performance noise" (Murphy 1997). In addition, the type of linguistic knowledge produced in online oral tasks may not always be a full reflection of the extent of knowledge that can be shown in comprehension tasks (Murphy 1997; Rothman 2007). In order to avoid overreliance on only one mode, two further tasks were also used (Task 2 and Task 3), based on White and Juffs (1998). As discussed in section 2.7.1 in the literature review (chapter 2), White and Juffs' study investigated acquisition of L2 English subjacency constraints among two groups of Chinese speakers of English: one with and the other without any immersion in English, and found native-like levels of accuracy among the non-immersion group. White and Juffs also found an object-subject asymmetry, and concluded that processing difficulties in subject extraction may underpin individual differences in overall scores of linguistic competence. White and Juffs' test methodology was adapted here to produce two pencil and paper tasks for this study, one for grammaticality judgements (Task 2) and one for written production data (Task 3), in order to gain additional insight into the potential systemic variation behind participants' question formation.

As noted in sections 2.7.4.3 and 2.6.6 in the literature review, concerns have been raised (e.g. Birdsong 1989; Bialystok 1994, 2002; Mandell 1999; Sorace 2003) about the conceptual and methodological validity of using grammaticality judgements as tapping implicit vs. metalinguistic knowledge. Grammaticality judgement tasks, as was shown in the literature review discussions, have long been used in generative linguistic research as a tool for native-speaker introspection of grammaticality, and were then adapted for early SLA research. However, when testing certain types of structures, such as past tense, or grammatical simple question forms, which form part of most instructed school curricula, disentangling metalinguistic knowledge of "grammar rules" from implicit competence of what is targetlike is difficult, if not impossible. There is also the issue that in reacting to often highly decontextualised prompts, respondents may be affected by the semantic plausibility of the tokens used, or simply guess. However, evidence from comparing grammaticality judgement tasks with other tests of linguistic knowledge (e.g. Mandell 1999) suggests that grammaticality judgement tasks are a reliable source of information about an L2 user's implicit linguistic knowledge. In addition, as a mechanism for investigating what cannot be taught because of ungrammaticality (such as subjacency violations), and provided that care is taken to minimize the chance of random guessing or semantic implausibility, grammaticality judgement tasks remain a valid means of linguistic data collection. Furthermore, since the design of this study relies closely on the design of other studies such as Juffs and Harrington (1995), and, in particular, White and Juffs (1998), which used grammaticality judgement tasks, I concluded that it was appropriate to use this methodology in this study.

Task 2: Grammaticality Judgement task

Task 2 (see Appendix A) asked for graded judgements of grammatical acceptability on twenty-two tokens of different complex question forms, derived from the party scene used in the first task. These consisted of twelve tokens of grammatical long-distance wh-movement (matched for subject and objects and finiteness and non-finiteness) and ten ungrammatical tokens (four subject tokens, four object tokens and two adjunct tokens). Ungrammaticality tested tense marking and subjacency: under or over-specification for tense (including do-support), and subjacency violations on adjunct extraction. Eight distracter tokens (four grammatical and four ungrammatical) were included to make a final task length of thirty items. Instructions were given on the paper, and two samples were given at the top of the paper to illustrate how to complete the task. Verbal instructions were also given to the whole group to check they understood the task. Examples of ungrammaticality of tense-marking and adjunct extraction are shown below.

(1)	*Who did Tom expect give the present?	Subject
(2)	*What did John know did Ann like?	Object
(3)	*What did Tom bring a present after he sent?	Adjunct

Participants indicated their judgement of the grammatical acceptability of the item on a six-point Likert scale of –3 to +3 (with 0 as a "don't know" option). Native-like judgements of –3 or 3 for ungrammatical or grammatical were scored as 1, to give a final accuracy score out of 22, although degree of variation was also recorded. This wide scalar range was intended to minimize the possibility of random guessing.

Task 3: Question production task

Task 3 (see Appendix A) again followed White and Juffs (1998), asking participants to generate written questions from a set of 15 answers where the target question word was underlined. These were balanced for subject and object extraction (grammatical and ungrammatical) and ungrammatical adjunct extraction and complex noun islands. Instructions and examples were given at the top of the paper and an oral check was carried out with the group before the start of the test to ensure the task had been understood.

An example of a given stimulus with a potential grammatically accurate (target-like) question is given in (4) below.

(4) A. Mary thought John enjoyed <u>the cake</u>.
Q. *Potential question: What did Mary think John enjoyed?*

In scoring the responses, White and Juffs' analysis was followed, differentiating between grammatically accurate questions that showed evidence of constraints on extraction, and those that did not, either because questions violated constraints on extraction, or avoided extraction, or were ungrammatical. Accurate responses were coded as 1, out of a possible total of 15. See examples (5) – (8) below from participants' responses:

(5) A. Ann knew <u>Tom</u> had sent her a card.
Q. Who did Ann know had sent her a card?

(6) A. Ann thanked John before she expected <u>Tom</u> to arrive.
Q. *Who did Ann thank John before she expected to arrive?

(7) A. Ann said that she was expecting <u>Tom</u>.

Q. Who was Ann expecting?

(8) A. Books about <u>Shakespeare</u> made Ann happy.

Q. *What books which made Ann happy?

Both these pencil and paper tests were carried out with no formal time constraints, although instructions were given to complete both as quickly and instinctively as possible, rather than spending time searching declarative memory or metalinguistic "rule knowledge" (Mackey and Gass 2005: 51), and time was called on each task after 10 minutes; all participants had finished both tests within that time. Although the tests may have seemed cognitively demanding (Skehan and Foster 1997; Robinson 2001) and potentially open to metalinguistic or controlled processing (Birdsong 1989; Bialystok 1994), the participants reported no difficulties in completing the tasks.

3.3. Working Memory tasks

Working memory data have commonly been gathered in L1 and L2 using versions of Daneman and Carpenter's (1980) seminal Reading Span Test, discussed in section 5.5 in chapter 2, the literature review. This test (RST) uses increasingly large sets of complex sentences, where participants read each sentence in the set, and then when prompted, try to recall as many of the sentence-final words as they can (Harrington and Sawyer 1992). However, a number of concerns arise from relying on this test in L2 studies, as discussed in the literature review (section 2.10). To recap these issues in brief, firstly, there are difficulties in finding robust correlations between RST and general measures of language proficiency (Juffs 2001; Sagarra 2000; Yoshimura 2001).

There are also issues of how far WM is language-independent or affected by any kind of bilingual language knowledge, whether at lower levels of proficiency or in balanced bilinguals switching between language mode (Grosjean 1982; Osaka and Osaka 1992; Service et al 2002). Secondly, research suggests that WM's role in linguistic performance appears to be task-specific (Waters and Caplan 1996, 2003), in that different modalities of the span task (reading, listening and speaking) correlated with language proficiency tasks within the same modality but not across different modalities, e.g. RST did not correlate with oral proficiency (Fortkamp 1999). Furthermore, WM tests have been focused on certain populations, e.g. highly educated university students. The validity of complicated tasks, such as RST, would therefore not be appropriate for certain L2 populations, such as beginner learners and low-educated learners (Juffs and Guillermo 2006), which undermined the universality of WM as a concept. In addition, WM tests, at least in studies I have found, have not moved on to investigate newer models of the WM construct, in particular Baddeley's episodic buffer, which, as discussed in the literature review, is theoretically suggested to act as a capacity-limited storage buffer for currently activated long-term linguistic knowledge, and has been implicated in the use of complex syntax in L1 and bilingual speech (Fehringer and Fry 2007). Finally, there have also been methodological issues identified from relying only on a single WM measure, and controversy over task-specificity (Waters and Caplan 1996; Conway et al 2005), allied to a growing interest in memory research into language used in "naturalistic" settings as a more appropriate measure than decontextualised tasks such as RST (Fry 2002).

These issues led to the question of how to test WM in a way that could be valid for all population types, using both L1 and L2 to control for language-dependent effects, and which used all the elements of Baddeley's WM model. I thus designed three WM tasks, one non-verbal: Digits Back, and two verbal: Listening Span and Story Recall. The Digits Back and Listening Span tests were designed to test the executive capacity of the WM model, using adaptations of standard WM tests. The aim of testing executive capacity in this way, rather than via a simple phonological short term memory (STM) storage task, was to test the dual-task element of WM, or more precisely the storage-processing trade off within WM, identified by Daneman and Carpenter (1980), which has been most strongly associated in the research literature with complex linguistic behaviour. The secondary purpose of using two tasks for this purpose was to compare non-verbal with verbal scores, to shed some light on the controversy over how task-specific WM measures may be. The Story Recall task was designed to test the construct of the episodic buffer, based on findings from L1 (Fry 2002) and bilingual studies (Fehringer and Fry 2007), as described in the literature review (chapter 2, section 2.6.5). All tasks were purely auditory, with no visual or paper stimulus for the participants, to simplify the procedure and make both the tasks and testing procedure as "naturalistic" as possible, while also avoiding reliance on reading strategies (Sagarra 2000). The tasks are summarised in the table below.

Table 4: Overview of WM tasks

Working Memory Tasks	**Focus**	**Procedure**	**Measure**
Digits Back L1	Executive capacity	Repeating strings of numbers in Mandarin of increasing length in reverse order. Two strings given per set.	Length of string when both strings of set correctly recalled (up to max of 7).
Digits Back L2	Executive capacity	As above but in English. Three strings given per set.	As above
Listening Span L2: Word Span Sentence Span	Phonological loop storage; executive capacity	Repeating sets of directions of increasing word length, then recalling specified words of all sentences in a set. Trial stopped if no words could be recalled.	Word span (phonological loop storage) measured longest sentence accurately recalled (out of 12). Sentence span (executive capacity) measured the longest sentence when specified words could be accurately recalled (out of 12).
Story Recall L1	Episodic buffer capacity	Listening to short narrative and immediately recalling as much as possible using same words and phrases.	Accuracy of gist and morphosyntactic recall (out of 50)
Story Recall L2	Episodic buffer capacity	As above but in English.	As above

3.3.1. WM Task 1 – Digits Back

The first task, Digits Back (Wechsler 1997) was chosen due to its wide use in L1 research, and also in L2 WM research (e.g. Walter 2003; Kormos and Safar 2008), thus acting as a reliable benchmark, easy to administer, and unrelated to linguistic proficiency (since advanced learners would all be familiar with the English digit names). More importantly, it would provide a crosslinguistic comparison between individuals' scores in Mandarin and English (in the light of documented differences between Digit Span scores in Mandarin and other languages, e.g. Chincotta and Underwood (1997).[9] This would shed further light into investigations as to how far WM is language-independent (Osaka and Osaka 1992) or affected by differences in L2 processing, and might also illuminate if any potential tendency to use L1 for counting, even by advanced L2 speakers, needed to be taken into account (Cook 1997; Service et al. 2002).

For Digits Back, participants heard sets of numbers read aloud at a rate of one digit per second (two strings per set in Mandarin and three strings in English). They were told at the start of the test that these sets would increase in length by 1 digit in each set up to a maximum of 7 digits. After each string, participants repeated the numbers in reverse order. Each new set was announced with the words, "Now for set 2, 3, 4 etc" as appropriate. The English version was read from a print out by the researcher; the Mandarin version was recorded as a digital sound file by a bilingual colleague who had been rated by three other non-tested Mandarin speakers from different areas in mainland China and Taiwan as having a particularly clear and neutral pronunciation.

The methodology here was slightly different from usual since the data was collected from the participants in a single group (see section 3.5 below). The usual procedure in individually administered tests is that only one string needs to be repeated correctly; one "second chance" is given for each length of string if the participant fails at the first attempt. However, all participants heard the whole set each time, to allow for any participants who required a second go. The sets were of a different size in English (three strings per set) to allow for lower proficiency/processing difficulties in the L2. Multiple strings per set also allowed for an alternative scoring procedure of "partial-credit scoring" (Conway et al 2005: 775), where participants' scores can reflect, in more valid and reliable ways, their performance across all elements of the task. However, in this first study, scoring was calculated following the more traditional "all or nothing" score using the length of the set where two strings were last recalled

[9] Chinese digits are monosyllabic and faster to articulate than English, resulting in higher average scores of Chinese WM measures of simple Digit Forward recall compared to English (Chincotta and Underwood 1997; Cheung et al 2000), but evidence of this difference has not, to my knowledge, been investigated for Digits Back, at least in L2A studies. Additionally, it has not, as far as I can establish, been identified whether any potential tendency to use L1 for counting, even by advanced L2 speakers, plays a role in L1-L2 differences.

correctly (Conway et al 2005: 774), in order to conform to more usual practice in using this test (Walter 2003).

3.3.2. Verbal WM tasks

As outlined above, the verbal tasks were created to address the issue of task-specificity and naturalistic language use, and to widen the investigation of WM to include the episodic buffer.

3.3.2.1. Task 2 – Listening Span

This test was devised in keeping with the focus on naturalistic WM tests, based from the concept of listening span (Daneman and Carpenter 1980; Harrington and Sawyer 1992), and derived from the idea of repeating directions (see below). The purpose of this test was to examine the offset between storage and processing through the combined testing of phonological loop short-term storage (word span) and central executive capacity (used as the usual Listening Span in WM research, and termed here as Sentence Span).

Three factors drove this design: one was anecdotal observation from my teaching experience of the difficulty L2 users can have in giving and understanding directions at all levels of proficiency (see Robinson 2001 for a discussion of the cognitive complexity of tasks involving direction-giving). The second was the aim of creating an easily-administered test that was rooted in an everyday use of WM. The third factor was the goal of devising a single test that deliberately targeted all the components of the Baddeley Model by including spatial language, and utilised both storage and processing.

Pairs of directions using the words "left", "right", "up" or "down" were created, increasing by 1 in each pair from five to twelve words (see Appendix A). The vocabulary used in this task was checked against the British National Corpus as being within the 1000 most common words, and created no difficulties in comprehension for any of the participants.

Participants were told that they would hear pairs of sentences containing one of the four specified direction words; the sentences would get increasingly longer (to a maximum of twelve words), which was designed to test participants' phonological loop capacity. They were to repeat each sentence aloud immediately, while remembering the direction word, which appeared in different places in each string to ensure the whole string would be listened to, testing their executive capacity. An example pair of sentences (seven words long) is given in (9) below.

(9) Walk **up** the street until the lights.
Take the second turn on the **left.**

Participants thus listened to, then repeated, the first string, then heard and repeated the second string; then after saying the second string, they tried to recall "up" and "left" in that order.

Scoring, like the Digits Back test, was a "quasi-absolute" score, measured as two scores. The first yielded a score for Word Span: the longest sentence length when the whole sentence was correctly recalled (out of a maximum of twelve). The second yielded a score for Sentence Span: the longest sentence length (again out of twelve) when direction words were correctly recalled. Due to limitations of time, this was only done in English (L2).

3.3.2.2. Task 3 - Story Recall

The final task specifically addressed the construct of Baddeley's episodic buffer. As discussed in my literature review, story recall tasks have long been used in L1 studies – my task is derived from the Story Recall sub-test from the Adult Memory and Information Processing Battery (AMIPB - Coughlan and Hollows 1985). The AMIPB task is used in L1 memory research as a test for schematic recall of both semantic and syntactic structure, particularly in the kind of naturalistic language use focused on here, and, as argued in my literature review (chapter 2), this task could therefore be an appropriate test for Baddeley's episodic buffer. This test had been specifically identified as correlating with use of complex syntax such as subordinate clauses and adverbial phrases in L1 (Fry 2002), and was theoretically of interest in looking at other types of complex syntax such as long-distance question forms as studied here. In addition, as far as I am aware, virtually no research on L2 WM using Story Recall has yet been published.[10] The purpose of the task in this study was thus firstly to assess whether it was a reliable and valid means of testing WM; secondly, to investigate if there were any correlations between Story Recall scores and the use of complex question formation in L2. I tested it in L1 and L2 to see if there were crosslinguistic issues of narrative recall which showed differences in schematic story construction (Rumelhart 1974).

Two WM Story Recall tests were devised in which participants listened to two stories, one in Mandarin (54 seconds long) and a different one in English (33 seconds), both based on the Coughlan and Hollows (1985) Story Recall test mentioned above (see Appendix A for the stories, both given in English). The first story, closely adapted from Coughlan and Hollows, was translated into Mandarin (L1) and recorded as a digital sound file by the bilingual colleague who had helped with the Digits Back task.

The second story in English (L2) had a similar schematic structure, but was shorter to avoid possible "floor effects" due to task difficulty (Harrington and Sawyer, 1992: 28).[11] The English version deliberately used complex syntax such as subordination, relative clauses and adverbial phrases to test the theoretical assumption made above that the episodic buffer storage/processing could facilitate accurate linguistic recall in complex syntax. The English version was read aloud by the researcher from a printout. In pilot tests, accuracy of recall ranged from 33% to 90%, suggesting the task level was appropriate.

Participants were asked to listen carefully to a short story which they were to repeat using the exact words and phrases as far as possible. The participants practised with three stories increasing in length from 11 words to 22 words. The actual story prompt was announced with a short instruction, "Now I'm going to tell you a story …". The scoring for both tests was adapted from the AMIPB scoring system, and consisted of ten points for recalling the ten narrative sections in order, and between two and three points for the syntactic and semantic elements in each section (such as connectives,

[10] Fehringer and Fry (2007) are, to my knowledge, the only study to test for correlations between Story Recall and language proficiency; I am grateful to the authors for permission to adapt their Story Recall task for this study.
[11] The L2 English Story Recall task was reviewed by Fry (p.c) and judged to conform suitably to the standard and structure of the original Coughlan and Hollows version.

verb forms and collocations). Participants could score up to a maximum of 50 points in total, which was then converted to an overall accuracy percentage score on each Story Recall test. Two bilingual colleagues were recruited to assess the Mandarin versions (one of whom was the translator and "voice" for the task). The raters were trained in the scoring system used in the English task, and worked with me in scoring the Mandarin versions to ensure scoring reliability.

3.4. Participants

3.4.1. Recruitment

Participants from mainland China and Taiwan, with Mandarin as L1, were recruited from a cohort of newly arrived postgraduates at British universities, between July and August 2006. All the participants had been in the UK less than two months at the time of first testing; all were previously classroom instructed learners. These recruitment specifications reduced the possible confounding effects of L1 transfer or variation in naturalistic exposure to ensure as much homogeneity as possible.

Participants were invited to participate in the research via the offer of two free group language development sessions, including vocabulary and cultural awareness, run one week apart in tandem with data collection at Time 1. The distraction activities were intended to divert participants from focusing on the study's research aims.

Forty Chinese and Taiwanese speakers signed up for the classes, but only eighteen students attended the first part of the initial two-part data collection session (see below), three of whom were excluded because they failed to return for the second part of the initial data collection session at Time 1 (yielding a participant group of 15 at Time 1). Four more did not return for further data collection at Time 2 (four months later) or Time 3 (around ten months after arrival), yielding a final set of valid longitudinal data for eleven participants.

3.4.2. Biodata

Bio-data on sex, country of origin and English proficiency level were obtained through completion of a questionnaire (Dörnyei 2005), which also contained questions to control for factors identified as affecting inter-learner variation in exposure to English (age of first exposure, length and type of exposure) as discussed in my literature review (e.g. Long 1996; DeKeyser 2000; Flege and Liu 2001; Moyer 2004). The questionnaire (see Appendix A) was laid out primarily as closed questions, in order to be easily understood and quickly completed in a group setting. All information was kept confidential and participants signed the questionnaire to indicate their permission to use the personal information in the data analysis.

The group consisted of three participants from Taiwan, and twelve from mainland China; there were six males and nine females. All participants were postgraduate students over the age of 20 who had achieved an IELTS score of between 5.5 and 7 in the previous four months. All participants therefore could be considered as advanced adult speakers of English, but not native-like (IELTS 9), and would still be expected to show variability in their morphosyntactic competence in question formation.

The age at which participants first started learning English (AOL) ranged from 10 to 15 years old (mean 12.07, SD 1.67), and the number of years they had been using English, including at school (LOL), ranged from 8 to 15 years (mean 11.73, SD 2.34). The hours of extra exposure to English outside school during their teenage years was measured: four of the 15 participants had had between 2 and 5 hours per week of naturalistic exposure, consisting mainly of watching English films or listening to English pop music.

Data were also collected on length of residence in the UK and amount of exposure in the UK (hours per week or month) to identify any potential effects of naturalistic exposure in the few weeks since participants had arrived in the UK (Dechert and Raupach 1987; Rothman and Iverson 2007).[12] In addition participants were asked if they had learnt a third language in order to allow for possible confounding L3 transfer effects (Leung 2005). Only one participant had any experience of an L3 (Japanese), and that was for less than one year, so this variable was excluded from further analysis.

SPSS analysis of correlations between IELTS scores and sex, country of origin, age of exposure, length and extra levels of exposure found no significant correlations. The

[12] Research on "study abroad" data suggests the first 3 weeks can be sufficient for a noticeable difference in accuracy in language production – Munoz p.c.). One participant (EM) had spent six months the previous year at a language school in the UK; however, since this participant's IELTS score and scores in the linguistic tasks were at or below the group average, the data are retained in the analysis.

group of participants was thus assumed to be as homogenous as possible in terms of exposure to English, and would therefore provide an appropriate group for researching potential correlations between linguistic proficiency and WM without major confounding factors arising from prior exposure.

3.5. Data Collection

As stated above, there were three instances of data collection, in line with the study's longitudinal design and research hypotheses. In order to test changes in wh-movement, linguistic data were collected at Time 1, within 2 months of arrival, Time 2 within six months of arrival, and Time 3, eleven to twelve months after arrival. WM was expected to remain stable and further testing of WM was thus not incorporated into the test design – scores from Time 1 would be used in conjuction with the longitudinal linguistic data, as in other development studies (e.g. Sagarra 2000). The study was initially designed to collect data in a group format, which would enable the data to be quickly and efficiently collected, and maintain parity throughout the data collection procedures; the group format design was also intended to yield a statistically valid and reliable number of participants (at least twenty in a single test collection session), to facilitate the repeated-measures quantitative analysis that would address my hypotheses most effectively, and with minimum burden of time or potential stress on the participants.

3.5.1. Time 1 (group format) – tasks and data collection procedure

At Time 1, there were two ninety-minute sessions which were conducted a week apart. Bio-data, oral and written data were collected during the first session, and the WM data collected in the second session. This separation of data collection was intended to avoid stress or harm to participants from test overload, and ensure that they did not get too tired or overburdened during data collection. It also was intended to avoid differences in WM if the tests were run on different days (Mattthews et al 2000).

The oral, written and WM data were gathered using the Sanako Lab 300 (Tandberg International) digital interactive software (version 6.0.0.1), delivered in a language laboratory in two universities which participated in this study. The software is designed to allow interaction between a central output (the teacher or researcher as live, or recorded digital source files) and up to 20 participants sitting at satellite computers, who can all hear the central output source, and whose own speech is captured as input in the form of a discrete digital sound file for each activity. The pause and switch function in the Sanako Lab Software allowed the tester to record the participants' individual responses to prompts, without recording the tester's intervention, thus reducing the size of the sound files (the maximum was around 3500 KB per participant). These groups of files are saved at the end of each activity in a central folder under the log-in ID of the computer each participant used, and were subsequently annotated with the participants' names after each test session. See below for the details on each task

Participants sat in front of computers, with headsets to allow individual communication with the tester. In the first session, they were assigned into pairs, and given distractor activities to learn new phrases, and practice these in getting to know each other, in line with what was offered at recruitment. This also gave them practice in listening to instructions through the headsets and talking to each other using the microphones on the headsets. The participants were then given the first linguistic task (see section 2.1 above) in which they were given paired sets of semi-completed pictures (as shown in Appendix A) and were given seven minutes to ask each other questions to find the differences between their pictures. The computer software captured each individual's speech through the headset microphone. The participants then removed the headsets to complete the second and third linguistic paper and pencil tasks individually, and complete the biodata questionnaires. They numbered their biodata questionnaires and linguistic task sheets with their computer ID, so that each participant's written data could be matched with the recorded data files.

In the second session, a week later, participants returned to the language laboratory. After a short session of language skills, practising general vocabulary, as a distractor task, the participants completed the Working Memory tasks (see section 3.3 above). As in the previous weeks, they listened through their headsets to instructions and the prompts.

Participants practised the Digits Back task with two non-recorded sets of three digits, repeating each string in reverse order. After each string was presented, the tester pressed a pause button on her console which allowed the software to switch to recording each participant's speech. There was a pause for a maximum of ten seconds between each string, or until all the participants had finished.

For the Directions task, the participants practised two pairs unrecorded, repeating each sentence immediately, and then recalling the specified direction words at the end of each set. As for the Digits Back task, each participant's response was recorded after each sentence was presented, with a maximum delay of ten seconds after each pair of sentences to allow for the word recall element of the task.

For Story Recall, they practised with three mini-stories increasing in length from one to three sentences, repeating what they heard as far as possible using the same words and sentences. After the actual test story was presented, participants took their own time to repeat it, and the recording session closed when all the participants had finished.

The tasks using English (both linguistic and WM) were presented orally by the tester through the central output system. The WM tasks in L1 Mandarin were recorded as digital sound files and loaded on to the Sanako Lab 300 software central source. The L1 tasks were done together at the end of the data collection session, to avoid constant changing between L1 and L2 during testing. The instructions on how to complete each task were pre-recorded as part of the WM sound files in both English (in the tester's voice) and Mandarin (by the colleague who recorded the WM prompts), and participants were reminded that the procedure would be the same as for the WM tasks they had already done in English.

3.5.2. Methodological difficulties

3.5.2.1. Problems with group format at Time 1

Despite the theoretical advantages of this group design for efficient, reliable data collection, there were a number of methodological and practical issues. Firstly, as highlighted above, the study was affected by participants' failure to return after Time 1 for subsequent data collection at Time 2 and Time 3, reducing the original pool of participants from eighteen to eleven, which affected the consistency and validity of the repeated measures tests, especially in the pairings for the oral task . Additional difficulties arose in maintaining the consistency of the tasks. The primary difficulty arose in the oral question elicitation task, where the group design meant that each pair of participants could adapt the pragmatic requirements of the task differently; some pairs followed instructions closely and only asked questions with very little "off-task" dialogue. Other pairs rapidly turned the task into a set of statements about their own picture, with few questions about their partner's picture. Others talked relatively little about the picture and used much of the time on "off-task" dialogue. The consistency of the task could not be maintained, and the results show this difference in task consistency.

Initial results from Time 1, summarised in the table below, show a wide range of individual variation. The total output in seven minutes for all questions (including all stages and whether accurate or not), ranged from 10 to 29, and for all utterances from 14 to 54. The Question Total (questions from stages 4 to 6) ranged from 4 to 17, but that included a large number of questions at stage 4 (copula fronting). Looking only at stage 5 and 6 questions, the range was 0 to 6 (mean, 2.36; SD, 1.88). Such small numbers of stage 4, 5 and 6 questions could undermine the validity of using these data as the source for defining the key measures of question total and question ratio.

The large SD especially for total utterances reflect the different approaches taken to the task across the group, raising concerns that other factors, such as pragmatic competence and personality (Dewaele and Furnham 1999), would have a measurable effect on the results (as an example of "performance noise"). These findings suggested that it would be difficult to reliably replicate the tasks using this group format at later data collection points (Time 2 and Time 3), undermining the validity of the repeated-measures longitudinal design.

Table 5: Summary of oral output at Time 1

	Mean	SD	Range
Question total (stages 4-6)	7.91	3.94	4-17
All questions (all stages)	16.91	6.76	10-29
All utterances	41.91	11.38	14-54

Secondly, as noted in section 2.10 in the literature review, WM data can be easily affected by affective external or internal variables such as time of day, hunger or tiredness (Matthews et al 2000; Fry 2002). By splitting the initial data collection into two parts, one week apart, the WM data ran the risk of being considered non-relevant, undermining the possibility of any correlation with linguistic performance.

3.5.2.2. Subsequent data collection

In order to minimise the difficulties arising from data collection at Time 1, I changed the language laboratory group format for subsequent data collection at Time 2 (six months after arrival), and Time 3 (around ten months), in favour of interviewing each participant individually. The face-to-face meetings made it easier to keep the participant "on-task", especially in the oral question elicitation task. The timings for each individual in this task were less consistent, as each dialogue had its own pragmatic constraints driving the total time used to complete the task (range at Time 2 was six to 10 minutes), but at Time 3, a maximum of 7 minutes was set at the start of the task, which meant the data would be comparable with the original data captured at Time 1 (on arrival). Oral and WM data from these meetings were collected using an Olympus WS-300 digital voice recorder, and a Sony ECM-MS907 condensing microphone, and stored as .WAV or .WMA files.

These subsequent data collection sessions used the same tasks, with slight amendments to the oral picture task and the grammaticality judgement task, to control for greater task efficiency, and avoid task memory or boredom effects. In the oral question task, there were fewer items on the participant's picture in order to stimulate more questions, and the grammaticality judgement task, in the second data collection session, used different distractor items. However, as the participants did not recognise the stimuli in this task, the original grammaticality judgement task version was used for the final test session at Time 3. The changed procedure also allowed all the WM and linguistic tasks to be completed in a single one-hour session, which was more convenient to the participants, and overcame the problem of variation in WM arising from affective factors. The same room was not always available, but it was always a quiet office or meeting room on the university campus to minimise environmental distractions.

3.6. Results and discussion

As outlined at the start of this chapter, I had four hypotheses, investigating asymmetries in acquisition of different question forms, the impact of immersion on development of these question forms, and the potential implication of WM on different question forms. I hypothesised that greater WM would correlate with faster progression from simple to complex question forms (i.e. greater use of the higher stages of question forms in Pienemann's hierarchy by Time 3), but that WM would not correlate with judgements on subjacency violations.

3.6.1. Oral data

Looking first at the linguistic data, and starting with the oral task, the raw data of question total and question ration at Time 1, Time 2 and Time 3 are presented, along with the degree of change over the period of immersion (calculated by subtracting scores at Time 3 from scores at Time 1). The question total shows the total number of questions from stages 4 to 6 in Pienemann's hierarchy of L2 English questions (copula questions, wh-questions using copula, and do-support; wh-questions using lexical verbs; and embedded questions). Question ratio shows the question total divided by the total number of utterances to show how far the participants used or avoided questions in the task. These two measures were designed to provide global measures of accurate production of increasingly complex question forms, in line with hypothesis 1 and 2. The results are shown in the two tables below:

Table 6: Oral question total from Time 1 to Time 3 (Mean, SD and Range)

	Time 1	Time 2	Time 3	Change by Time 3
Mean	7.91	13.91	9.36	1.45
Std. Deviation	3.94	7.20	5.10	7.22
Minimum	4	5	3	-14
Maximum	17	25	21	10

Table 7: Oral question ratio from Time 1 to Time 3 (Mean, SD and Range)

	Time 1	Time 2	Time 3	Change by Time 3
Mean	.22	.30	.17	-.04
Std. Deviation	.14	.14	.08	.17
Minimum	.10	.09	.09	-.36
Maximum	.50	.55	.39	.09

The results show that individual variation in the oral task was pronounced, with a wide range of scores for question total at Time 1 and Time 3, and for the overall change over the period of immersion. Descriptively, the mean question total slightly increased by Time 3, but the mean ratio slightly decreased. Using Wilcoxon signed-rank analysis, there were no significant differences between Time 1 and Time 3 for either total or ratio. Looking at the degree of change, five of the group showed zero or negative change in question total (from 0 down to -14) and question ratio (0 to -.36); in other words nearly half the group showed no positive improvement in asking questions as the year went on. The reduction in totals was mainly due to a decrease in stage 4 (copula questions); there was no overall change in stage 5 or 6 questions.

Non-target forms from the Time 3 data include primarily omission of verbal morphosyntax in all question types, and some over-suppliance. Examples produced by different participants are given in (10) to (18) below, with individual participant IDs shown in capitals in brackets:

(10) What was Tom study? (HER)

(11) What he drinking? (HER)

(12) Has any drink on the table? (CHI)

(13) How is he look like? (IVY)

(14) Do you have information about where Tom come from? (MAY)

(15) Who prepare for the snack? (CHE)

(16) Is the music starts? (CHE)

(17) Where does the music from? (CHE)

(18) What's are in the bowl on the table? (HOR)

The implication from the oral data was that development in question formation remained highly variable even at the end of nine months' immersion, and that head movement (or tense marking) rather than wh-movement per se was the primary difficulty. It emerged from informal post-hoc interviews that all the participants felt they were no better at speaking than when they had arrived in the UK, and seven of the group judged that they had got noticeably worse during the year. The participants felt they could understand spoken English (either informal conversation or academic lectures) better then when they had first arrived in the UK, but did not feel more communicatively or pragmatically fluent or competent, unlike other studies which have shown improvements in fluency but not accuracy during time spent studying abroad (e.g. Freed 1995).

However, rather than conclude that these data mean that there was little evidence of acquisition of underlying abstract features, it is possible that the data reveal processing difficulties at the morphology/syntax interface, in line with the Missing Surface Inflection Hypothesis (Prévost and White 2000; Lardiere 2007), since the main sources of non-targetlikeness for the participants lay in morphological tense marking. It could also be that the lack of significant progress as measured reflects a change in processing different sources of linguistic information, (in line with the coalition model suggested by Herschensohn 1999), or different levels of activation of underlying features (suggested by Truscott and Sharwood Smith 2004, in press), or the beginnings of reanalysis of learned chunks (Myles 2004). In other words, the data could provide evidence of some kind of "tipping point" at which reliance on explicit learned knowledge starts to reduce in favour of using implicit (but as yet not fully specified) competence. These speculations could be borne out by evidence from the paper tasks, which tested a wider range of question forms, testing in more detail for the asymmetries predicted in my literature review.

3.6.2. Pencil and Paper data

I now present the data from the pencil and paper tasks to see if the observed lack of change in the oral task was also evident in these tasks, as speculated above. Given the lack of significant difference between Time 1 and Time 3 in the oral data, I present the data just for Time 1 and Time 3, without the Time 2 data or change data, for ease of presentation (full tables including Time 2 and change data are given in Tables i and ii in Appendix C).

The pencil and paper tasks were the grammaticality judgement task and question formation task. These were designed to give more specific information than could be elicited in the more spontaneous oral task, in order to test for asymmetries between taught forms and untaught forms (including subjacency violations), and between subject and object questions. The grammaticality judgement (GJ) task focused on subject/object asymmetry, comparing taught grammatical versus untaught ungrammatical complex question forms, but included two tokens on adjunct extraction; the question formation (QF) task focused on subjacency violations including extraction from complex NPs and adjunct violations, but included grammatical long-distance extraction. The GJ task was marked out of 22; the QF task out of 15. Evidence of targetlike accuracy on ungrammatical forms in the GJ task and on the subjacency violations on the QF task could be taken as evidence of acquisition in terms of implicit competence, since knowledge that these tokens were ungrammatical could not be part of learned explicit knowledge. The results are summarized in the tables below.

Table 8: Grammaticality Judgement data

(out of 22)	Time 1	Time 3
Mean	8.36	10.45(*)
Std. Deviation	3.32	3.26
Minimum	4	5
Maximum	16	16

(= significant difference found between Time 1 and Time 3, p<.05)*

Table 9: Asymmetries in Grammaticality Judgement task between grammatical and ungrammatical questions

	Grammatical questions (/10)		Ungrammatical questions (/10)	
	Time 1	Time 3	Time 1	Time 3
Mean	3.91	5.82	4.45	4.73
Std. Deviation	2.59	1.72	2.50	2.49
Minimum	0	3	2	1
Maximum	8	9	8	10

Table 10: Asymmetries in Grammaticality Judgement task between object and subject questions

	Object questions (/10)		Subject questions (/10)	
	Time 1	Time 3	Time 1	Time 3
Mean	5.18(*)	5.82(*)	3.27	2.73
Std. Deviation	1.54	3.12	2.45	2.054
Minimum	3	0	0	1
Maximum	8	9	8	7

(= significant difference found between objects and subjects at both times, p<.05)*

Table 11: Question Formation data

(out of 15)	Time 1	Time 3
Mean	7.55	10.36
Std. Deviation	3.96	2.94
Minimum	1	6
Maximum	13	15

The results of these two tasks showed higher levels of morphosyntactic accuracy than in the oral task, as could be expected given the lower levels of "online performance stress" in paper format (Murphy 1997), but accuracy was still far from target-like at Time 3, even after more than nine months' immersion. At Time 3, maximum accuracy in the grammaticality judgement (GJ) task overall was 16 out of 22, or 72% (achieved by one participant), while the mean for the group (10.36 out of 22) was only 47%, or around chance. However, there was significant improvement in mean scores between Time 1 and Time 3, according to Wilcoxon signed rank analysis (p<.05).

In terms of the expected asymmetries tested in the GJ task between taught and untaught forms (grammatical versus ungrammatical), the expected greater accuracy in grammatical forms was not found. Mean accuracy in ungrammatical forms was descriptively higher than for grammatical forms at Time 1, but lower at Time 3, but this difference was not significant (p>.2). There was greater improvement in grammatical forms than ungrammatical forms, which approached statistical significance (p=.082). Object questions were significantly more accurate than subject forms at both Time 1 and Time 3 (p<.05). Descriptively, subject forms were judged less accurately at Time 3 than at Time 1, although the differences between Time 1 and Time 3 were not significant for either subject decrease or object increase.

Scores in the Question Formation (QF) task showed greater accuracy, with maximum scores of 15 or 100% achieved by one participant; this was not the same individual as the highest scorer in the GJ task. Five of the group scored above 70% (11 or above). The mean score for the QF task descriptively shows there was improvement from 7.55 (50.3%) to 10.36 (69.1%), which showed a trend towards significance, according to Wilcoxon signed rank analysis (p=.09). Since the QF test was specifically designed to test sensitivity to subjacency constraints, the higher overall accuracy in this task suggests that wh-movement in its most abstract syntactic sense had been acquired by Time 3 (using Vainikka and Young-Scholten's (1994) 60% targetlike accuracy as acquisition).

These data, taken together with the higher accuracy scores on ungrammatical forms in the GJ task, point to the Missing Surface Inflection Hypothesis (Prévost and White 2000; Lardiere 2007) as being a reasonable account for the tense marking difficulties found in the oral task. However, given the lack of statistical significance, this conclusion remains somewhat speculative. The speculation raised above, that the variability could indicate some kind of "tipping point" where different sources of knowledge were being utilized, could not be robustly confirmed. There were tentative indications of some kind of reliance on implicit knowledge even at Time 1 (given the higher scores on the ungrammatical or untaught tokens in the GJ task, and good scores on the QF task), but counter to that, the accuracy on the grammatical or taught forms in the GJ task improved significantly more than ungrammatical forms over the period of immersion, suggesting that reliance on explicit knowledge remained an important element of L2 processing.

In addition, there were also interesting indications of the different test modes tapping similar bases of linguistic knowledge shown in some significant positive correlations between the different linguistic tasks at Time 1 and Time 3, using Spearman non-parametric correlation; significant correlations are shown in the tables below.

Table 12: Correlations between linguistic tasks at Time 1 and Time 3

	Question total (Time 1/Time 3	Question ratio (Time 1/Time 3)	GJ (Time 1/Time 3)	QF (Time 1/Time 3)
Question total Time 1	-	-	-	
Question total Time 3	-	.765(**)	-	.614(*)
Question ratio Time 1	.805(**)	-	-	-
Question ratio Time 3	.765(**)	-	-	.805(**)
GJ Time 1	-	-	-	-
GJ Time 3	-	-	-	-
QF Time 1	-	-	-	-
QF Time 3	.614(*)	-	-	-

(** Correlation is significant, $p<.01$; * Correlation is significant, $p< .05$)

The GJ measure showed no significant correlation with any of the other measures, but oral question total, question ratio and QF scores showed highly significant correlations with each other, supporting the tentative indication of some kind of reliance on implicit knowledge being established by Time 3. The lack of correlation between the GJ scores and QF scores remains hard to explain, especially since, theoretically at least, the GJ task and QF tasks were intended to map onto related types of complex question formation.

The data thus remain inconclusive in terms of confirming or disconfirming my hypotheses. None of the participants could be said to have acquired target-like morphosyntax across all modes of testing, taking a standard of 60% accuracy in obligatory contexts as acquisition. There were a number of reasons for using this criterion of acquisition – the use of percentage accuracy was consistent with White & Juffs (1998), and in addition, allowed consistent comparison across all 3 tests, and with other acquisition literature using triangulated data collection methodologies, whereas using a measure such as Pienemann's emergence criterion would be limited to the oral data only.

In terms of patterns of variability and asymmetry, complex sentences and subjacency violations showed some degree of target-likeness, but the expected difference between taught and untaught forms was not found; object questions were, however, found to be easier than subject questions. The awareness of tense-marking required in wh-movement shown in the written tasks was not revealed in the oral task, which had high levels of non-targetlikeness in omission of do-support and verbal inflection. In all, it seems that immersion appeared to facilitate improvement of linguistic knowledge, especially in the GJ and QF tasks, but was not sufficient to trigger statistically significant restructuring of underlying linguistic knowledge towards targetlike levels of accuracy evident in all modes of language use.

3.6.3. Working Memory correlations

Given the lack of significant improvement in linguistic accuracy over the period of immersion, using individual variation in linguistic development as the basis for correlations with WM scores was potentially going to be problematic. My research hypotheses had been that WM capacity was implicated in these learners' ability to access existing knowledge of taught question forms more efficiently, in that greater WM capacity would correlate with individual differences in rates of increasing targetlike use of complex questions, and that WM capacity was not implicated in their capacity to acquire untaught implicit subjacency constraints. These hypotheses would be supported if positive correlations were found between WM (tested at Time 1) with linguistic accuracy at Time 1, subject to the asymmetries tested above, and with degree of change between Time 1 and Time 3.

To test these hypotheses, the data were analysed using non-parametric correlational analysis using SPSS. The first set of correlations, between WM and changes in linguistic tasks over the period of immersion, shown in the table below, show no significant correlations between WM and change in any of the linguistic tasks.

This is in line with other recent studies looking at WM and L2 morphosyntactic development (Sagarra 2000) and WM and L2 fluency (Mizera 2006) which did not find significant correlations.

Table 13: Correlations between WM data and degree of change in linguistic tasks

	DBL1	DBL2	SRL1	SRL2	Wordspan	Sentencespan
Change Oral Question total	-.066	-.146	.079	-.299	.007	.224
Sig (2-tailed):	.846	.687	.828	.372	.984	.507
Change Oral Question ratio	-.046	-.024	-.171	-.422	.183	.434
Sig (2-tailed):	.894	.947	.637	.196	.589	.183
Change GJ	.253	-.340	-.078	-.019	-.040	-.151
Sig (2-tailed):	.453	.337	.831	.957	.908	.658
Change QF	.000	.086	-.006	-.395	-.133	.053
Sig (2-tailed):	1.000	.813	.987	.229	.698	.876

To try to find secondary evidence that WM was not implicated in individual differences in linguistic accuracy, correlations were also run on linguistic scores at Time 1 and Time 3. In the oral task, for Question total, at either time, there were no consistent trends and all results were nonsignificant ($p>.29$), in line with the results shown above. For Question ratio, some clearer trends emerged. All WM tasks correlated positively with question ratio at Time 1 and Time 3, with some trends towards significance. At Time 1, Story Recall approached significance (for L1, $r=.589$, $p=.073$, for L2, $r= .543$, $p=.084$); at Time 3, Sentence Span approached significance ($r=.561$, $p=.073$). The correlations are shown in full in Table iii in Appendix C.

For the paper tasks, the trend overall was for positive correlations, as shown in the table below, although very little trend towards significance was found. Interestingly, however, Sentence Span, argued to be the closest to the standard Listening Span Test, showed significant positive correlation with the QF task at Time 3 ($r=.629$, $p<.05$), argued to be the test most focused on implicit untaught knowledge (of subjacency constraints).

Table 14: Correlations between WM and GJ and QF tasks overall scores

	DBL1	DBL2	SRL1	SRL2	Wordspan	Sentencespan
GJ task Time 1	.350	.394	.320	.279	.419	.451
Sig (2-tailed):	.292	.259	.368	.406	.200	.163
GJ task Time 3	.303	.198	.099	.153	.269	.298
Sig (2-tailed):	.366	.583	.785	.653	.424	.374
QF task Time 1	.346	.099	.380	.390	.415	.570
Sig (2-tailed):	.297	.786	.279	.236	.204	.067
QF task Time 3	.148	.262	.357	-.132	.293	.629(*)
Sig (2-tailed):	.664	.465	.312	.699	.383	.038

(* Correlation is significant, p<.05)

Looking in more detail at the asymmetries in the GJ task, (grammatical versus ungrammatical forms, and subject versus objects), there was no clear evidence of correlation between the linguistic and WM scores (full correlations are shown in Table iv in Appendix C). There were no overall positive or negative trends and most results were nonsignificant (p>0.1) at either Time 1 or Time 3, or with changes in accuracy over the period of immersion. There was a pattern of positive but non-significant correlations between all WM tasks and ungrammatical questions at Time 3; Sentence Span showed moderate near-significant correlation (r=.540, p=.086). These data taken in conjunction with the positive correlations found with the question formation task and the Sentence Span task seem to counter-indicate the presumption at the start of my study that WM would correlate with explicit taught knowledge but not with untaught knowledge such as subjacency violations or ungrammatical forms.

3.7. Evaluation of first study

In general terms, therefore, there was no robust evidence of significant correlations, or of very clear patterns in general in the data. This, I argue, could primarily be down to the methodological flaws already discussed above (small group numbers, participant dropout, group versus individual data collection, the low numbers of suitable questions elicited in the oral task). It was also pointed out (Juffs, Fry, p.c.) that the design of the Listening task, which was intended to combine a phonological loop storage measure (Word Span) and an executive capacity measure (Sentence Span), could be conceptually flawed, since these two factors were not usually intended to be tested in combination in most standard psychological WM tests.

These flaws rendered the data from this first study to some extent invalid and unreliable. It was therefore important, in redesigning the study, to establish firstly whether linguistic development could be more reliably measured using revised tasks, and secondly, whether the assumption that WM could be the key to L2 variation, as Miyake and Friedman (1998) claimed, could still have some test validity, in order to be retained as a hypothesis in the main study.

For the first assumption about linguistic development during immersion, the findings seemed to suggest a complex interaction between different types of linguistic knowledge which may arise due to performance or processing issues – i.e. participants could recognise head movement and subjacency constraints offline in paper-based tasks, when there was no time pressure, but they could not produce questions using head movement as accurately in spontaneous online speech. These findings seemed compatible with the notion of task-specific demands on processing (e.g. Robinson 2001; Bialystok 2002), in which different tasks require different processing strategies or different sources of knowledge, such as metalinguistic awareness for the paper tasks (Bialystok 1994).The data are also compatible with White and Juffs' (1998) conclusion that differences in processing rather than in competence might be key to explaining variation in L2 acquisition, in that "L2 learners might achieve similar competence to native speakers, and yet take longer to access that competence" (White and Juffs 1998: 127). This needed further detailed examination in ways that could tap more precisely into language processing, such as using a reaction time computerized grammaticality judgement task, rather than the paper-based tasks used here. The additional information garnered on reaction times would provide useful comparison with other studies using timed tasks, particularly those testing automaticity (e.g. DeKeyser 1997; Segalowitz 2000; Juffs 2004; Sunderman and Kroll 2009). The inconclusive nature of linguistic development also needed further investigation, to see if the overall conclusion that a year's immersion would not necessarily lead to longitudinal improvement was reliable.

For the second assumption about the role of WM in SLA, it was shown that there were some positive correlations, approaching significance, between WM tests and certain elements of the linguistic tasks, e.g. question ratio and question formation. However, the evidence of significant or near-significant correlations between WM and implicit knowledge seemed to counter-indicate the assumption at the start of this study that greater WM capacity would favour the retrieval and improved use of explicit taught knowledge. The speculation that WM could be the key to L2 variation, as Miyake and Friedman (1998) had claimed, seemed to be more indicative of growing accuracy in implicit knowledge, rather than explicit knowledge. These findings needed further careful investigation using a wider range of explicit and implicit forms, which were better balanced between grammatical and ungrammatical or taught and untaught types than had been used in the paper tasks used in the first study. It was also clear that the WM tasks themselves required careful revision and changes in how they were to be used, needing modifications to the listening task design, and to ensure WM tasks were tested at the same time as linguistic tasks all through the study design, in order to make correlational analysis more valid.

The findings therefore confirmed that studying cognitive factors affecting L2 acquisition such as WM remained a viable research theme; moreover, a key question about the role of WM in accessing different types of linguistic knowledge under pressure remained unanswered. A second, more extensive, study was therefore designed to focus more precisely on this area, and also incorporated changes to revise the recruitment and testing procedures carefully to take account of the methodological and test design flaws discussed above. The second study will be discussed in the following chapter.

Chapter 4: Second study- revised hypotheses and methodology

4.1 Introduction

Although inconclusive, the results of the first study described in the previous chapter nonetheless showed that there remained a rationale for investigating the role of WM in individual variation in second language acquisition, seen from positive and significant or nearly-significant correlations between WM capacity and improvements in both spoken and written tasks. The main conceptual thrust of the original assumptions remained unchanged, but a key methodological adaptation was to change the pencil-and-paper grammaticality judgement task into a more extensive computerised reaction time task. The bi-modal concept of Schwartz (1993) and Ullman (2001) still underpinned the investigation, and the additional use of a reaction time task was aimed to clarify whether sources of linguistic knowledge used in pressured situations (such as speech and online timed tasks) were primarily either explicit or implicit, or some "coalition" of both (Herschensohn 1999).

Miyake and Friedman's (1998) claim that WM would be implicated in individual variation in rates of development was retained, in order to explore how far this related to forms reflecting explicit taught knowledge but not implicit knowledge. The use of a timed task would also provide interesting data to check for potential correlations between greater WM and reaction times (Kroll et al 2002; Sunderman and Kroll 2009). The first study had revealed some differences in accuracy between object and subject questions (Schachter and Yip 1990; White and Juffs 1998), and I was interested to discover how far this difference was also reflected in processing speeds (Juffs and Harrington 1995). The assumption of immersion affecting linguistic development was also retained, but extended to apply to both greater accuracy and faster processing in the reaction time task, to compare against the lack of robust evidence of significant improvement found in the first study.

In addition, there were many methodological problems with the first study, outlined in the previous chapter, including small sample size, inconsistent test procedures (group versus individual data collection) and format of some of the linguistic and WM tasks. In line with these revised hypotheses and in order to improve on the flaws, there were some procedural revisions to the data collection design, and the linguistic and WM tasks were redrafted. The three key revisions were changing from a group format to individual face-to-face data collection throughout the study, changing the paper tasks to a timed computerized grammaticality judgement task, and revising the Listening WM task.

The research hypotheses were thus revised for a second study as follows:

Hypothesis 1: Instructed Chinese learners of English will show asymmetries in acquisition and use of question forms, tested through oral output and timed grammaticality judgments, measured as:

a) greater use of targetlike simple questions compared to complex questions in oral output
b) faster speed and accuracy in timed judgments on simple questions than complex questions, especially untaught implicit forms, including subjacency violations
c) faster speed and accuracy in timed judgments on grammatical questions compared to ungrammatical questions and object questions compared to subject questions.

Hypothesis 2: These learners will improve over time in their knowledge and use of question forms (subject to the asymmetries noted in 1) when they are exposed to enriched input through increased primary linguistic data from native speakers in an immersion setting.

Hypothesis 3: WM capacity is implicated in their ability to access existing knowledge of taught question forms more efficiently, in that greater WM capacity would correlate with individual differences in perception/production of targetlike questions (other than for subjacency-constrained items) on both linguistic tasks, and with variation in rates of improvement over time during immersion.

Hypothesis 4: WM capacity is not implicated in their capacity to acquire untaught implicit subjacency constraints, measured in a timed grammaticality judgement task.

4.2. Tasks

The tasks for the second study were closely based on the materials used for the first study, with some key changes, which are described below. All revised materials for the second study are given in Appendix B, and all tasks are summarized in the table below. All testing was carried out in an individual interview with each participant, in a quiet room on university campus or equivalent, using a digital sound recorder and microphone to capture the data, as used in the individual data collection sessions in the first study.

Table 15: Second study task summary

Linguistic Data	What tested	Measures	*Differences to first study?*
Oral data: Question elicitation picture task	Accuracy and fluency in spontaneous oral output	Accuracy (total number of target-like stage 4 and 5 qus) Fluency (ratio of qus. to utterances, 0-1)	*No change (see first study, ch. 3, section 2.1.1)*
Grammaticality judgement data: Reaction Time task	Accuracy and speed in timed sentence processing	Speed in ms; Accuracy (target-like responses, max: 68)	*Changed (see Appendix B)*
Digits Back in L1 and L2	Non-verbal executive capacity	Accurate recall of increasing string length (max: 7) scored as ratio (0-1)	*No change in task, change in scoring (see ch 3, section 3.1)*
Story Recall in L1 and L2	Verbal executive and episodic buffer capacity	Accurate recall of meaning and form (max: 50)	*No change (see ch 3, section 3.2.2)*
Listening Span	Verbal executive capacity	Accurate recall of pre-specified words (max: 5) scored as ratio (0-1)	*Different design (see Appendix B)*

4.2.1. Oral task

The oral question elicitation task remained the same as in the first study, although the timing was confined to five minutes for all participants at both times of testing, to ensure greater task consistency. The form of the picture used to elicit questions was that used in the later individual interviews in the first study (e.g. at Time 3), with minimal information on the participant's picture, in order to reduce avoidance and elicit as many questions as possible. The participants were given clear instructions that they were to ask as many questions as possible using wh-question words, in order to gain information about the scene indicated in the picture, so that their picture would match the tester's picture as closely as possible.

The face-to-face interview format ensured that interactive constraints were consistent, in order to reduce the socio-pragmatic difficulties that could have affected the reliability of data from Time 1 in the first task. The processing element was strengthened by a time constraint (five minutes only) and the instruction to ask as many questions as possible in that time. The context of the picture (a party scene with four characters) was designed partly as a way of creating a context for the stimuli used in the second linguistic task, a timed grammaticality judgement task. This was to maximize the semantic consistency of the two tasks, and diminish decontextualisation.

Scoring for the oral task, as in the first study, consisted of a measure for accuracy, by counting the total number of target-like questions showing head movement and wh-movement (represented by Pienemann's stages 4-6: copula questions and use with wh-expressions; yes-no questions using do-support, auxiliary inversion, and wh-questions using lexical verbs; embedded questions). As before, a secondary score was question ratio, taking the total number of utterances in the time given, and calculating the proportion of questions per utterances, to provide a measure of fluency. These two scores for accuracy and fluency provided the baseline against which the WM measures could be tested for correlation with accurate speech production in line with the first part of research hypothesis 1.

4.2.2. Reaction Time Task

The indications of positive correlations between oral and paper tasks found in the first study suggested that using the oral task again and retaining some kind of grammaticality judgement measure was justified, maintaining the connection with White and Juffs' (1998) methodology for my original study design. However, given the signs of interesting information about how language knowledge is retrieved and processed under pressured situations such as speech compared to paper-based tasks, as well as evidence of asymmetries between grammatical and ungrammatical forms, and subject and object questions, it was clear, as discussed in the previous chapter, that there were limitations with using untimed paper-based tasks for such grammaticality judgements.

Recent research into sentence processing has stressed the usefulness of using online processing, especially through measuring reaction times to stimuli, to gain greater insight into the complexities of L2 acquisition (Juffs and Harrington 1995; Murphy 1997; Marinis 2003; McDonald 2006). E-prime and PsyScope have been among the most commonly used (Marinis 2003: 154). However, very recent research in L1 language processing has used DMDX software (Forster and Forster 2003), which is an easily accessible and quickly learnt alternative, and is available as downloadable freeware.[13]

For my second study, the original set of complex questions used in the grammaticality judgement task and question formation task in the first study were pulled together, and added short-distance questions to create a data file consisting of sets of simple and complex questions. These stimuli tested short movement (set at Pienemann's stages 5 and 6), balanced for type and word length, and also long movement, including subjacency violations, closely following White and Juffs (1998) and Juffs and Harrington (1995); they were as far as possible matched for grammaticality versus ungrammaticality (thirty-eight of each) and subjects and objects (twenty-eight of each). Also included were eight items testing resumptive pronouns and eight examples of tense omission acting as distracters and ensuring an equal number of grammatical and ungrammatical items in the overall test. Discounting the distracters, there were sixty-eight test items. The full set of items is attached in Appendix B, but examples are given below.

For the simple short-distance questions, twenty grammatical subject and object questions were matched with twenty ungrammatical questions, which omitted or over-produced one morphosyntactic element, either tense marking or do-support.

Example stimuli are given below:

(1) (a) *Who was arrive later by car? (subject)

(b) Who was arriving later by car?

[13] Version 3.1.6.2 (2006) was used in this study. DMDX software was developed by Ken and Jonathan Forster at Monash University and the University of Arizona. Available at http://www.u.arizona.edu/~jforster/dmdx.htm

(2) (a) *What was Tom eat with his fingers? (object)

(b) What was Mary eating with her fingers?

(3) (a) *Do you see was who eating cake? (subject)

(b) Do you know who was eating cake?

(4) (a) *Do you have information about where is Mary come from? (object)

(b) Do you have information about where Mary comes from?

For the complex long-distance questions, twenty-eight stimuli targeting subjacency violations were adapted from from Juffs and Harrington (1995), testing ungrammatical subject islands, relative clause islands, adjunct islands (twelve items), and lexically matched grammatical finite and infinitival subject and object extraction (sixteen items). Examples are given below:

(5)	*What did books about make Ann happy?	Subject islands
(6)	*What did Mary see the man who stole?	Relative clause islands
(7)	*Who did Ann thank John after she saw?	Adjunct islands
(8)	Who did Ann say liked her friend?	Subject extraction (finite) [14]
(9)	Who did Mary say her friend liked?	Object extraction (finite)
(10)	Who did Ann want to win the game?	Subject (non-finite)
(11)	What did John want to win?	Object (non-finite)

These stimuli were converted via DMDX into individual items, presented on a laptop computer screen. Each participant was presented with an introductory screen asking them to make a judgement about some sentences they were about to see, if they were "grammatically acceptable" or not, using a four-way Likert scale scale of +2 (acceptable) to –2 (unacceptable), activated by pressing specially marked buttons at either side of the keyboard, as shown in the figure below. Studies consulted to inform this study commonly used a simple binary choice to show acceptance or rejection (e.g. Juffs and Harrington 1995; Sunderman and Kroll 2009), but I wished to retain some breadth of scale in order to minimize guessing, and ensure that information was

[14] In more current generative theory the long-distance infinitival grammatical questions (10 and 11) would be termed as containing PRO clauses, containing null subject pronouns (Radford (2004: 108-110), but for consistency with Juffs and Harrington's study I use their terminology of finite/non-finite.

obtained on grades of accuracy as well as reaction time (Sorace 2003). It could also provide data on potential changes from lack of certainty to greater certainty over time, and how such changes might indicate development of more targetlike or systematic automatised linguistic knowledge. However, the scale was reduced from the six-way scale used in the paper task to a four-way scale, in order to be more practicable. In addition, the fine-nuanced six-way division used in the first study did not produce any significant or marked information on changes in participants' intuitions over time, so a four-way scale was decided to be appropriate.

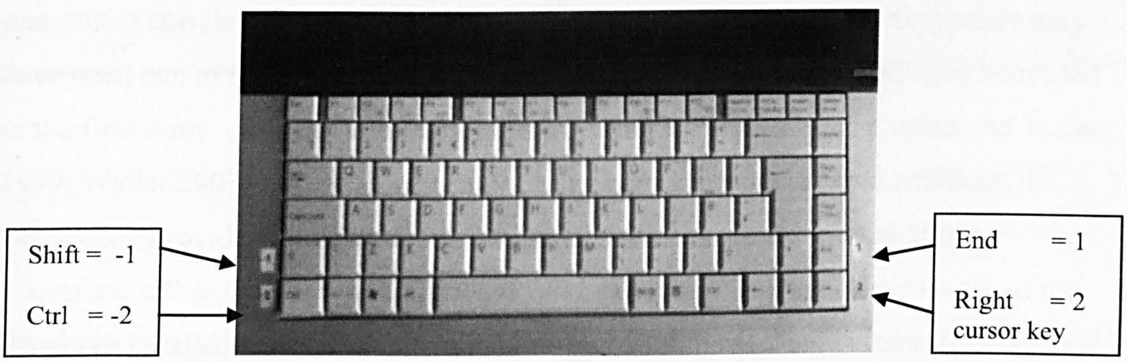


Figure 6: Keyboard marked for Reaction Time task

Each participant was instructed prior to the test to respond as quickly as possible with their first intuitive response to each stimulus, according to a sense of "grammatically acceptable" or not. Care was taken to ensure that the specific phrase of "grammatically acceptable" was understood (as advised in Juffs and Harrington 1995), so that participants knew they should respond without too much mental searching for possible metalinguistic grammatical knowledge of taught rules (Mackey and Gass 2005: 51).

The stimuli appeared in full, centred on the screen in Comic Sans font size 24.[15] The task was self-timed: pressing the space bar started the clock which continued until one of the buttons was pressed; this action also then generated the next stimulus to appear, in a different random order generated by the software for each participant. The test continued until the words "Test finished – Thank you" appeared on the screen accompanied by a beep. The data could then be saved as a single digital file in SPSS-

[15] As advised by Miller and Letts (p.c. Speech Department, Newcastle University).

compatible format by pressing the ESC button. The whole test was then restarted when needed for the next participant. Three practice stimuli items were presented to ensure the participants understood how to use the buttons appropriately.

4.2.3. Working Memory tasks

4.2.3.1. Digits Back in L1 and L2

This task was identical to the task used in the first study: participants heard strings of digits in pairs (in Mandarin) and in sets of three (in English); the larger sets in English were to offset greater potential processing difficulties, as explained in the previous chapter (section 3.1). The only difference was in the scoring; the lack of any obvious patterns of correlation between linguistic data and Digits Back in the first study may have been due to the absolute scoring method used. The absolute scoring system used in the first study, although common in both L1 and L2 studies (e.g. Caplan and Waters 1999; Walter 2003), can produce heavy clustering over modal scores which do not necessarily provide valid data on individual differences (Conway et al 2005). Therefore, rather than use the traditional "all or nothing" scoring of the length of the string last recalled correctly (up to a maximum of 7), the scoring system in the second study followed Conway et al (2005)'s recommendation to use the partial-credit scoring system, which calculates an individual's mean score as a proportion of the correct number of digits recalled out of the possible maximum. The scoring system gives a maximum of 1 per correct string, and a partial score of the number of correct digits recalled divided by the maximum possible in that string (e.g. from a set of three strings of four digits, where the first and second string were correctly recalled, and the third string had digits recalled correctly, the mean score would be calculated as 1 + 1 + .5 (2/4). Each set would be calculated in this way, and then divided by the total number of digits recalled (up to a maximum of 12), to give a ratio or decimal score between 0 and 1. This system of partial credit scoring also creates a more reliable way of comparing WM scores when multiple different types of task are used (Conway et al 2005), as in the study design used here.

4.2.3.2. Story Recall in L1 and L2

Task design, instructions and scoring were exactly the same as in the first study. Participants heard a story in Mandarin and another story in English adapted from the AMIPB (Coughlan and Hollows 1985), and had to repeat them as far as possible using the same words and phrases. This innovative task, as outlined in section 3.2.2 in the previous chapter, was created in order to test the validity of the construct of the episodic buffer and its potential operation as a central focus to store and process existing and novel information (Baddeley 2000). The requirement to recall accurately meaning and form, targeting past tense marking and complex syntax such as relative clauses and subordination, provided information on how automatically these morphosyntactic forms could be retrieved, while focusing on overall gist, and had been found to correlate with use of complex grammar in L1 and bilingual L2 (Fehringer and Fry 2007). Some tentative indications of the validity of this task were evident in the first study, which found positive correlations, approaching significance, with accuracy in the oral task (for L1, $r=.589$, $p=.073$; for L2, $r= .543$, $p=.084$), and a clear trend of positive correlations with accuracy on the paper tasks, and the task was therefore retained for the second study. As before, responses were marked out of 50, consisting of up to 10 points for recalling up to 10 schematic elements in story meaning, and up to 40 points for accuracy in repeating the semantic and syntactic elements.

4.2.3.3. Listening Span task in L2

From the WM data, it was clear from the first study that there was some potential validity for assuming a connection between Listening Span-type measures of WM and linguistic proficiency, particularly in the Sentence Span measure. Sentence Span approached significance ($r=.561$, $p=.073$) with question ratio in the oral task at Time 3, and showed significant or near-significant positive correlations with the question formation task at both times (at Time 1, $r=.570$, $p=.067$; at Time 3, $r=. 629$, $p<.05$).

However, as outlined in section 6.3 in the previous chapter, concerns had been raised (Juffs, Fry, personal communications) that the combined design which tested both phonological loop storage and executive storage/processing did not follow standard WM testing procedure, which typically separates phonological loop measures from executive measures, and this could make the task, to some extent, invalid.

The Listening Span was therefore redrafted, still using directions, but closer in methodological design to existing types of listening span tasks which test recall of specified words over increasingly longer sets of sentences (Daneman and Carpenter 1980; Harrington and Sawyer 1992; see also methodological review in Conway et al

2005). Word frequencies were tested against the British National Corpus to reduce the question of low frequency words interfering with the aural processing.

The new Listening Span task consisted of increasing sets of sentences, all between seven and nine syllables long (eight syllables was the mean score of string length from the first test, and therefore taken as a suitable average length). Participants listened to each sentence and immediately repeated the direction given (e.g. "left", "right"), then at the end of each set they were prompted to recall the final words of both sentences in the correct order.

An example is given below:

(12)	Turn left after the train station	LEFT (repeated immediately)
	Go right at the supermarket	RIGHT (ditto)

Then "station" and "supermarket" were repeated when prompted.

Responses (out of a possible 27 in raw form) were scored following Conway et al (2005)'s partial credit scoring scheme, where the number of correctly recalled final words were divided by the maximum number of strings tested to give a ratio or decimal score between 0 and 1.

4.3. Participants

Given the potential confounding effect of the small group size in the first study, and the effects of participant drop out, a larger group of participants was recruited (n=40) from presessional English courses for international students run at the same British universities as in the first study. All were Mandarin L1 postgraduates from mainland China and Taiwan. Each participant was seen individually in quiet rooms either on campus or in the researcher's home. After completing the biodata questionnaire, the oral question task and the WM tasks, they were asked to complete the computerised Reaction Time experiment. Each individual test time took around 45 minutes and all the tasks were completed in a single session for each participant. As described above, the RT experiment was recorded using DMDX software installed on a laptop, so that the task could be administered in any location. The oral tasks were recorded using the same recording equipment as before (an Olympus digital recorder with microphone).

The biodata were gathered through a semi-structured interview with the researcher based on the questionnaire used in the first study, assuring the confidentiality of the data and gaining the permission of the participants. The questionnaire had been revised (see Appendix B) to include an added question checking if there had been any break in learning since school prior to the participant's arrival in the UK, to see if this would be an additional factor affecting retrieval of learned knowledge, perhaps in terms of less efficient processing or attrition, or lower levels of activation of L2 knowledge since the original period of instruction (e.g. Köpke et al 2007; Sharwood Smith and Truscott, ms). The interview mode of asking the questions ensured the questions were understood clearly and the responses were consistent.

The RT task and WM tasks were administered within the first three months of arrival (Time 1) and again between nine to twelve months of immersion (Time 2). The study used two data collection points, rather than three, as in the first study, because of the larger group size, potential difficulties with retaining participants and, most importantly, the lack of significant trends in linguistic development shown over the three sessions in the first study. However, the full period of a year's immersion was retained, to maximise the opportunity to compare data from the first study with the data from the second study, and to maximise the likelihood of finding some evidence of change. Difficulties remained with participant retention, and eight of the original forty did not return for re-testing at Time 2, leaving a final data pool of thirty-two participants. All the biodata and test scores were tabulated into SPSS for analysis.

The biodata were checked for key factors known to affect variation in exposure (as explained in the previous chapter): sex, country, age of learning (AOL), length of learning (LOL), additional exposure during formal instruction at school, break in learning before arrival in UK (and if so, how long). The principal findings are summarised in the table below.

Table 16: Summary of biodata

All participants (N=32)	N	Mean	SD	Min	Max
Female	24				
Male	8				
Taiwan	18				
People's Republic of China (PRC)	14				
AOL (years)		11.41	1.58	7	14
LOL (years)		11.77	2.91	5	18
Extra exposure at school - <2 hrs /wk	17				
>2 hrs /wk	15				
Participants with a break in learning (length in years shown in brackets)	11	(6.04	5.14)	(3	19)

The group was reasonably matched for country of origin, although there was a preponderance of females to males (the original group of forty contained twelve males but four dropped out at Time 2). The age range for when participants started learning English was 7 to 14 years (mean: 11.41 years). Length of learning showed a wide range, from 5 to 18 years (mean: 11.77 years), and sixteen participants had had a break in using English before arriving in the UK, ranging from 3 to 19 years (mean: 5.81 years). The group was equally split between those who had more than two hours a week additional exposure outside school and those with less. Two of the group had started learning prior to 10 years of age, but neither they nor any of the rest of the group had had exposure to naturalistic interactive native-speaker input, and so these two participants were retained in the group since they fitted the recruitment criterion of being instructed learners.

As in the first study, IELTS scores for all participants were also noted: all had achieved scores of at least 5.5 in recent IELTS tests, so that they were all considered to be at advanced level. Fourteen had achieved level 5.5, sixteen had achieved 6 and two had achieved level 6.5.

ANOVA tests showed no significant effects for sex or country on IELTS score, age or length of learning, additional exposure outside school or break in learning. ANOVAs were also run to see if age or length of learning showed any effect by group based on IELTS score or break in learning. Length of learning was the only factor that showed some effect. The two participants at IELTS level 6.5 were significantly different to those at level 5.5 in length of learning ($p<.05$); these two participants also had not had a break in using English since leaving school. The sixteen participants who had had a break in using English were significantly different in length of learning to those who had not ($p=.000$). In other words, the two participants who scored 6.5 in the IELTS tests before arriving in UK had spent longer learning English than those who had scored 5.5, and those who had had a break in using English prior to coming to the UK had spent less time learning English overall than those who had not. However, apart from these data, there were no other significant effects of age or length of learning or additional exposure outside school.

With the provisos in mind for the impact on subsequent data results of the individuals identified above who were potential extreme cases for age of learning or IELTS level, and for the potential effect of break in learning, I nevertheless assumed that there were no outliers prior to testing that did not fit the recruitment requirements for the study. Therefore in terms of effect of external factors arising from education history or amount of exposure prior to arrival, the participants were assumed to be as homogenous as possible.

The following chapter provides the data gathered in the linguistic and WM tasks to test the four hypotheses of this study.

Chapter 5: Second study - data results

5.1 Introduction

This chapter presents the data collected in my second study to test the research hypotheses investigating patterns of variation in development of L2 English wh-movement and correlations between these patterns of variation and Working Memory (WM). The first study, on a small scale, had identified asymmetries in question formation in oral and written tasks, but had revealed no clear trend of correlation between the results of these tasks and WM. In addition, the first study indicated that the expectation of significant linguistic development during the period of a year was not robustly supported. However, the first study was affected by a number of methodological issues, including limited numbers of participants (N=eleven) and issues of reliability in the written tasks, specifically the use of untimed grammaticality judgements, which meant the data were hard to interpret (see section 6.2 and 7, in chapter 3, detailing the first study).

These issues prompted revisions which drove the second study discussed here (see chapter 3 for details). The primary change was to revise the written task into a timed grammaticality judgement task, which covered a wider set of question forms, in order to test for evidence of asymmetries in question formation with greater effectiveness. The oral task remained the same and most of the WM tasks were identical, apart from the Listening Task (see section 2.3.3 of the previous chapter for details). Data were collected longitudinally over the period of one year, as in the first study, and using two points of comparison, Time 1 (within four weeks of arrival in the UK) and Time 2 (around eleven months later).

The revised hypotheses for the second study are explained and justified in full at the start of the previous chapter, but are summarised again here for ease of reference:

Hypothesis 1: Instructed Chinese learners of English will show asymmetries in acquisition and use of question forms, tested through oral output and timed grammaticality judgments, measured as:

a) greater use of targetlike simple questions compared to complex questions in oral output
b) faster speed and accuracy in timed judgments on simple questions than complex questions, especially untaught implicit forms, including subjacency violations
c) faster speed and accuracy in timed judgments on grammatical questions compared to ungrammatical questions and object questions compared to subject questions.

Hypothesis 2: These learners will improve over time in their knowledge and use of question forms (subject to the asymmetries noted in 1) when they are exposed to enriched input through increased primary linguistic data from native speakers in an immersion setting.

Hypothesis 3: WM capacity is implicated in their ability to access existing knowledge of taught question forms more efficiently, in that greater WM capacity would correlate with individual differences in perception/production of targetlike questions (other than for subjacency-constrained items) on both linguistic tasks, and with variation in rates of improvement over time during immersion.

Hypothesis 4: WM capacity is not implicated in their capacity to acquire untaught implicit subjacency constraints, measured in a timed grammaticality judgement task.

5.2. Data collection overview

In order to test these hypotheses, tasks were used based on or adapted from the first study, described in full in chapter 3, (section 2.1.1 and section 3), but summarised with revisions shown here for ease of reference.

5.2.1. Oral Measures

For oral data, a question-elicitation task was used to measure scores on four elements. The two principal scoring measures were:
Question Total - the number of target-like accurate questions asked in a 5 minute period (at levels four and five of Pienemann's hierarchy of processability, representing questions that showed targetlike verb-raising);

Question Ratio - the proportion of such questions out of the total number of utterances produced, shown as a proportion between 0 and 1, as evidence of task performance or avoidance.

Two further analyses were undertaken:
Complex forms - number of accurate indirect questions, tag questions and use of subordinate clauses, as evidence of complex sentence structure (Pienemann 1998, Young-Scholten et al 2005). This was to allow for comparison with predicted asymmetries in the reaction time task between simple and complex forms;

Nontarget forms - total number of non-repaired forms representing earlier stages in the implicational hierarchy of question formation (e.g. chunks, verbs used in-situ). In view of the participants' observed tendency to self-correct and monitor their speech in the first study, this measure was introduced in the second study to provide additional information on the balance of targetlike question forms compared to non-targetlike forms, to provide as complete a picture as possible on participants' underlying linguistic knowledge and their ability to produce targetlike structures under the online pressure of the oral task.

5.2.2. Reaction Time Measures

This new task was designed to garner more processing information on participants' processing accuracy and speed by using a computerised reaction-time grammaticality judgement task, rather than a paper-based task as in the first study. As detailed in the previous chapter (chapter 4), and laid out in full in Appendix B, the task consisted of 68 test items and 16 distractors, testing knowledge of wh-movement across three different wh-question structures: short movement, long movement, subjacency-constrained forms. The test was balanced between grammatical and ungrammatical forms (34 of each), and short and long movement tokens were balanced for subject and object questions (28 of each). Examples of each of the three wh-structures are given below:

(1) Short movement: What did John eat? (grammatical, object)

(2) Long movement: Who did Mary say liked her friend? (grammatical, subject)

(3) Subjacency-constrained: *What did the book about please Ann? (ungrammatical, complex NP extraction)

Analysis was run on participants' speed (RT) and targetlike accuracy as a total score for the whole task, then further analysis was carried out to compare RT and accuracy on grammatical and ungrammatical forms, and on each of the three wh-question structures under investigation. This subdivision by type was designed to test the hypotheses outlined above, that there would be significant asymmetries across different syntactic types, even after immersion.

5.2.3. WM Measures

WM data were collected following similar measures to those used in the first study (apart from the revisions made to the Listening task, as discussed in the previous chapter). As detailed in chapter 3, there were five tasks: the Listening Span task (in L2 English), Digits Back in L1 (Chinese) and L2 (English), Story Recall in L1 and L2. The Listening Span task and the Digits Back tasks which recalled increasing numbers of sentence-final words (up to a maximum of 4) or strings of digits (up to a maximum of 7) were scored following Conway et al (2005) as partial-credit scoring, taking all answers into consideration, as a proportion of the total number of items recalled. These tests resulted in ratio scores between 0 to 1 (see chapter 4, section 4.2.3.1 for scoring procedure in full). The Story Recall tasks were scored by accuracy of recall of both gist and grammatical phrasing out of a possible maximum of 50. However, for ease of comparison, the Story Recall tasks are also presented here as ratio scores. The mean and SD scores for all the tasks at both Time 1 and Time 2 were calculated to provide a baseline of how the different tasks compared to each other and across time to check for any unexpected patterns that could affect the subsequent correlations. The baseline analysis would also provide data to compare L1 and L2 versions of the Digits Back and Story Recall tasks to test whether WM was language-independent or not (Osaka and Osaka 1992; Service et al. 2002), and how far the various measures changed over time, testing the assumption in the L1 literature that WM is stable in adults between the ages of around 20 to 50 years (Baddeley et al. 2009).

Correlations between WM scores and the oral and RT data were run to test Hypotheses 3 and 4 above, i.e. that WM would be implicated in increased use of targetlike oral questions, especially use of complex questions, and would also be implicated in asymmetric ways with the RT data, in that that positive correlations would be found on explicit taught forms, such as short movement and grammatical items, but not with implicit nontaught forms, i.e. Subjacency-constrained items.[16]

5.2.4. Statistical analysis of normality

Statistical tests of normality of distribution were carried out on the linguistic and WM scores to test for the presence of outliers and non-normal distribution. Shapiro-Wilks tests showed significantly non-normal distribution across a number of variables ($p<.05$), and boxplot analysis revealed a number of outliers lying more than -/+ 1SD beyond the mean in the linguistic data. Given the mixed pattern of distribution it was assumed that the homogeneity of the whole pool of participants may be statistically affected. Therefore all statistical analysis for this study was carried out using non-parametric tests: paired-samples Wilcoxon signed rank tests were used to test for significant differences between scores at Time 1 and Time 2, and Spearman correlations were used to test for correlations between linguistic and WM scores. Analysis is first carried out on the whole group or pool of participants, and then, where relevant to explore the data further, the whole pool is divided into low, mid and high-performing groupings according to scores at Time 1 for each task, for between-group analysis using Kruskal-Wallis tests.[17]

5.3. Data presentation

The following sections present the linguistic data in detail to look at changes during the period of the study and to test whether the research hypotheses outlined at the start of this chapter were supported. The remainder of this section provides descriptive and inferential statistic analysis of the linguistic data showing results at both times of testing and the longitudinal change in results by Time 2. For ease and clarity, the oral data is presented first (section 5.3.1), and then the RT data (section 5.3.2), and in each section, the data is assessed to show how far the relevant asymmetries predicted in Hypotheses 1 and 2 regarding linguistic development are supported. Section 5.3.3 provides a brief overview of both sets of linguistic data to summarise the patterns of linguistic

---

[16] Due to technical difficulties with three individuals' sound files for some of the WM tests at Time 1, there are some missing scores which show in the different N sizes for the correlations (n=29 at Time 1, n=32 at Time 2).

[17] Each of these groupings is derived separately for each task or measure under discussion: e.g. the groupings on the oral task are not the same as the groupings on the RT task, and the groupings on RT speed are not the same as the groupings on RT accuracy.

development that were found. Section 5.4 provides descriptive statistics of the WM data at both times of testing; section 5.5 analyses the correlations drawn between the linguistic data and the WM data to test how far Hypotheses 3 and 4 about the predicted implication of WM in certain syntactic forms are supported.

5.3.1. Oral Data

This section analyses the oral data to assess whether targetlike question production showed significant asymmetries as predicted (Hypothesis 1), and whether targetlike question production changed significantly during immersion (Hypothesis 2).

5.3.1.1. Overall scores

As outlined above, there were four measures arising from an oral question-elicitation task at Time 1 and at Time 2, as follows: Question total, Question ratio, Complex forms, Nontarget forms. The predicted pattern of development in comparing Time 1 and Time 2 was that the score showing total number of accurate questions showing targetlike verb raising and wh-fronting would increase, and the ratio of accurate questions as part of the overall discourse would increase, that the total number of complex forms would increase and that nontarget forms (without verb raising) would decrease. The two tables below summarises the main descriptive statistics for Time 1 and Time 2, providing Mean, SD, Minimum and Maximum.

Table 17: Oral question task: descriptive statistics at Time 1

	Mean	SD	Minimum	Maximum
Question total	11.5	4.6	4	25
Question ratio	.27	.13	.13	.72
Complex forms	2.5	1.9	0	7
Nontarget forms	5.7	3.1	0	12

Table 18: Oral question task: descriptive statistics at Time 2

	Mean	SD	Minimum	Maximum
Question total Time 2	10.7	5.0	4	26
Question ratio Time 2	.30	.14	.11	.67
Complex forms Time 2	1.7	1.6	0	5
Nontarget forms Time 2	5.6	3.6	0	15

Figure 7 shows the direct comparison of the mean scores at Time 1 and Time 2. Table 19 summarises the change in scores between the two times of testing.

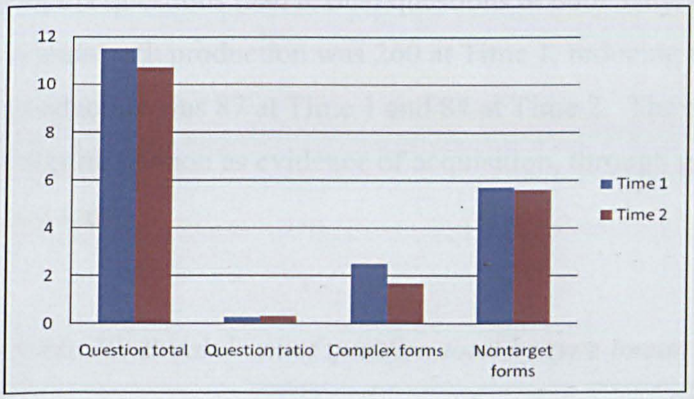


Figure 7: Mean scores in oral task at Time 1 and Time 2

Table 19: Change in scores between Time 1 and Time 2

	Mean	SD	Minimum	Maximum
Change in question total	-.53	5.12	-11	14
Change in question ratio	.04	0.152	-.23	.39
Change in complex forms	-.75	1.98	-5	3
Change in nontarget forms	-.13	3.60	-6	8

As can be seen from the data above, the group mean scores showed little change between Time 1 and Time 2. Using Wilcoxon signed rank analysis, the scores at both times were compared for significant change. All scores remained virtually the same, apart from Question ratio, which showed a slight improvement (non-significant), and nontarget forms, which showed a slight negative change (in other words, slightly fewer errors, again non-significant). Simple questions were significantly preferred to complex forms (p=.000), with some participants scoring 0 for complex forms at both times. Complex forms showed the greatest decrease between Time 1 and Time 2, and this was the only change which was statistically significant (p<.05).

In order to assess whether there were differential patterns of change by question type, the question total scores were re-analysed in line with Pienemann's implicational hierarchy comparing use of stage four questions (using copula) and stage five questions (showing lexical verb raising). Comparing the two verb types revealed a marked difference in usage, shown in the table below and illustrated in Figure 8 below. According to Wilcoxon signed rank analysis, there was a significantly greater use of copula questions than lexical questions at both Time 1 and Time 2 (p=.000). Total copula verb production was 260 at Time 1, reducing to 226 at Time 2; lexical verb production was 87 at Time 1 and 84 at Time 2. The expected increase in lexical verbs after immersion as evidence of acquisition, through gains in later stage questions, was not found.

Table 20: Breakdown of question total by type (mean scores)

	Mean	SD	Minimum	Maximum
Copula Time 1	8.35	3.33	2	17
Copula Time 2	7.63	3.71	1	18
Lexical Time 1	2.88	2.03	0	8
Lexical Time 2	2.88	2.47	0	11
Change copula	-0.72	3.91	-8	11
Change lexical	0.0	2.42	-4	7

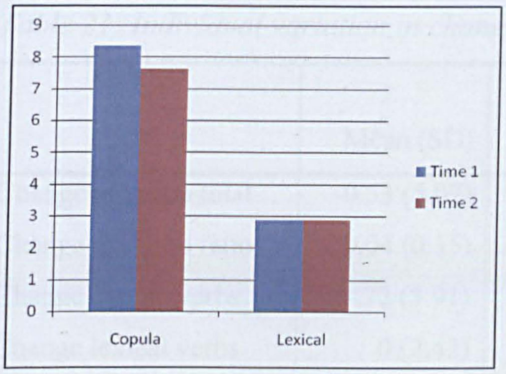


Figure 8: Mean scores broken down by question type at Time 1 and Time 2

The heavy bias towards use of inflected copula questions (three times as many produced overall as lexical verbs) may be due to a task effect arising from a strategy for checking or describing the picture, such as "Why is..", "Where is …", "What is…". Given the prevalence of singular "Is" rather than other possible forms of inflected copula such as "Are…?" or "Was…?" it is assumed that many of these could have been chunked lexical items (Myles 2004) or evidence of earlier stages of acquisition (Pienemann 1998), especially stage 3, which features verbs marked for tense but left in situ. If so, the use of chunks and early stage verb in-situ use remained high even at Time 2, again reinforcing the conclusion that the expected progress arising from immersion did not occur.

The results of overall mean scores thus show little significant change by Time 2; however, these overall scores derived from the whole pool of participants could obscure patterns of wide individual variation as indicated by high standard deviation values and wide differences between minimum and maximum scores. In other words, the descriptive statistics showed that a few individuals were scoring either very high or very low on all the measures being analysed, and it was possible that these individuals were affecting the overall mean scores.

In order to check this possibility, the change scores of the whole pool of participants (n=32) were analysed for extreme cases who lay more than 1SD (+/-) beyond the mean. The data in the table below show the group mean score (with SD in brackets), the number of all participants who decreased and increased over time (out of 32), and those extreme cases who decreased or increased beyond 1SD.

Table 21: Individual variation in changes in oral measures

	Mean (SD)	Total decreased	Total increased	Decreased below 1SD	Increased above 1SD
Change question total	-0.53 (5.07)	18	12	4	4
Change question ratio	0.04 (0.15)	14	18	3	5
Change copula verbs	-.72 (3.91)	18	12	4	4
Change lexical verbs	0 (2.42)	14	11	9	5
Change complex forms	-.75 (1.98)	14	18	3	5
Change nontarget forms	-.13 (3.61)	14	11	9	5

The data for individual variation in changes on the oral task confirmed the group scores analysed above, showing a greater number of participants whose scores decreased on nearly all measures, but that this change was within 1SD of the mean. In terms of extreme cases, between eight and fourteen participants showed change beyond 1SD of the mean; one participant was found among the five highest scorers across all the change measures, and one among the five lowest scorers across four measures, but there were no other individuals who scored among the extreme cases more than twice, suggesting that there was no clear pattern of a very few consistently high-scoring or low-scoring individuals who could be affecting the overall picture of little change.

For most of the measures, both group and extreme case totals showed a decrease, confirming the overall mean scores analysed above. However, for change in complex forms, the individual scores of change contradicted the negative group trend, since more than half the group actually showed an improvement. For change in nontarget forms, the group score showing only a very slight improvement again obscured the fact that more than a quarter of the group (nine participants) showed more than 1SD improvement (defined as decrease in nontarget forms).

In descriptive terms, therefore, it is possible to see that group mean scores suggesting little or no change across the oral measures obscured more marked individual patterns in certain cases (increase in complex forms and decrease in nontarget forms). However, it was unclear from the above analysis whether the decrease came from high scoring individuals at Time 1 who did not maintain their original high scores at Time 2, or from mid or low scoring individuals at Time 1.

5.3.1.2. Between-group analysis

In order to clarify this question statistically, the whole pool of participants was divided into three groups, split by equal percentiles, according to their score in question total at Time 1.[18] A Kruskal-Wallis test was conducted to see if the groups performed significantly differently to each other, in terms of question type (copula, lexical, complex, nontarget) at Time 1 and Time 2, and in change by Time 2. The Low and Mid groups contained twelve participants each, and the High group contained eight participants.

Time 1 results showed between-group differences for copula questions (chi-square $=14.160$, $p<.01$) and lexical questions (chi-square $=7.039$, $p<.01$), but no between-group differences for complex or nontarget forms ($p>.7$). There were no significant differences between groups at Time 2 ($p>.1$ on all measures), nor in the scores of change by Time 2 ($p>.1$ on all measures).

To clarify these differences, the data for copula and lexical questions, complex and nontarget forms were reanalysed to show between-group scores, and compared using Wilcoxon signed rank analysis for significant differences between Time 1 and Time 2, as shown in the tables and figure below.

[18] A four-way group split by -/+1 SD and mean created unbalanced group sizes, hence the equal percentile cutpoints were used to ensure comparability.

Table 22: Mean scores by group for each question type at Time 1

Group	Question type	Mean	SD	Minimum	Maximum
Low	Copula	6.18	2.10	2	9
	Lexical	1.58	1.51	0	5
	Complex	2.29	1.76	0	5
	Nontarget	5.14	2.88	1	9
Mid	Copula	8.00	1.91	3	10
	Lexical	3.08	1.31	1	5
	Complex	2.50	2.54	0	7
	Nontarget	6.08	2.78	2	10
High	Copula	12.13	3.44	8	17
	Lexical	4.50	2.45	1	8
	Complex	2.75	1.04	1	4
	Nontarget	5.88	4.02	0	12

Table 23: Mean scores by group for each question type at Time 2

Group	Question type	Mean	SD	Minimum	Maximum
Low	Copula Time 2	6.00	2.41	1	10
	Lexical Time 2	1.92	1.44	0	4
	Complex Time 2	1.50	1.83	0	5
	Nontarget Time 2	4.92	2.54	0	10
Mid	Copula Time 2	8.58	3.85	4	18
	Lexical Time 2	3.67	3.50	0	11
	Complex Time 2	1.67	1.37	0	4
	Nontarget Time 2	5.75	4.45	1	15
High	Copula Time 2	8.63	4.63	2	15
	Lexical Time 2	3.13	1.36	1	5
	Complex Time 2	1.88	1.64	0	5
	Nontarget Time 2	6.25	3.85	2	11

Table 24: Change in scores by group

		Mean	SD	Minimum	Maximum
Low	Change Copula	-0.17	3.49	-5	6
	Change Lexical questions	0.33	2.10	-3	3
	Change Complex	-0.67	1.97	-3	2
	Change Nontarget	-0.25	3.44	-6	5
Mid	Change Copula	0.58	3.96	-4	11
	Change Lexical questions	0.58	2.81	-2	7
	Change Complex	-0.75	2.22	-5	2
	Change Nontarget	-0.33	3.96	-6	8
High	Change Copula	-3.50	3.42	-8	2
	Change Lexical questions	-1.38	1.92	-4	0
	Change Complex	-0.88	1.89	-3	3
	Change Nontarget	0.38	3.74	-6	7

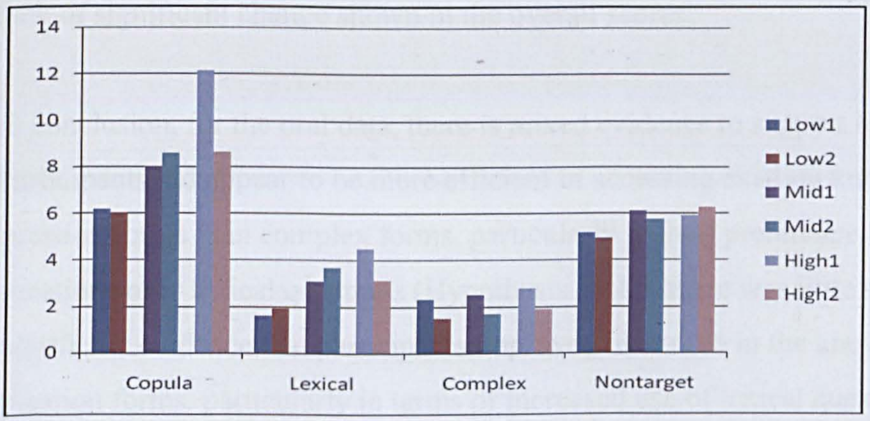


Figure 9: Mean scores across question type by group comparing Time 1 and Time 2

In this analysis of comparison across question type split by group, the groups remained statistically similar across both Time 1 and Time 2. The High group produced significantly fewer copula questions at Time 2 ($p<.05$). There were no other significant between-group differences across the other measures. Descriptively, the Low group improved slightly in lexical verbs, but decreased slightly on copula and complex questions. The Mid group improved in copula and lexical questions, but decreased in complex forms. The High group by contrast decreased in lexical questions at Time 2, producing fewer than the Mid group (in addition to the significant decrease in copula questions). While the Low and Mid groups improved on nontarget production (the Low group produced the least nontarget forms at both times), the High group's nontarget score went up at Time 2. However, these differences were slight, reflected in the lack of statistical significance.

Therefore it is clear that the pattern of oral output remained similar at Time 1 and Time 2, with little significant findings for change by question type or group level. The between-group analysis reflected the overall findings in the overall mean scores, suggesting that the wide range of individual scores did not have a significant effect on the overall scores.

Improvement in the oral task was clearest in the question ratio measure, and to some extent in improvements in lexical questions. The overall decrease in group mean question total (11.5 to 10.7) was identified as deriving from a decrease by the High group producing fewer questions at Time 2, especially copula questions. The Mid and Low groups showed no significant difference by question type by Time 2, reflecting the lack of significant change shown in the overall scores.

In conclusion, for the oral data, there is mixed evidence to support Hypotheses 1 and 2. Participants did appear to be more efficient in accessing existing knowledge of simple question forms than complex forms, particularly in their preference for copula questions over lexical questions (Hypothesis 1), but there was little statistically significant evidence that participants improved over time in the knowledge and use of question forms, particularly in terms of increased use of lexical questions, increased use of complex forms and reduction in nontarget forms, as a result of increased exposure in an immersion setting (Hypothesis 2).

5.3.1.3. Post-hoc microanalysis for patterns of change in oral data

In view of the unexpected lack of change in the oral data according to the research hypotheses, I carried out two post-hoc microanalyses to see what patterns of variation were evident in the data, looking at changes in nontarget forms and measures of fluency.

Firstly looking at nontarget forms, these were analysed to see if there were any identifiable predictable patterns to what participants found most difficult. As mentioned in section 5.3.1, the total number of accurate questions was around 11 or 10 at both times of measuring. The mean number of non-target forms, at 5.7 (Time 1) or 5.6 (Time 2), was around half this number.

Closer investigation of the types of non-target forms revealed the majority of problems arose from lack of targetlike verbal inflection marking. Most nontarget forms showed either omission or oversuppliance of morphosyntactic elements, primarily relating to *do*-support and auxiliary *be*, as well as mis-marking of number, e.g.:

1. When is the party start? (ALL)
2. Does she lived alone? (ANG)
3. Why he seems very unhappy? (CAT)
4. And do three guys er <is to coming> is in the house? (ALL)
5. What did Mary said? (CHI)
6. Is these two boys sitting on the chair or? (ALL)
7. Does Annie angry? (ELA)
8. How many people late for this party? (ROV)

Participants frequently demonstrated optionality (Sorace 2003) between targetlike and nontargetlike production, including lack of verb raising (leaving tensed verbs in-situ), and omission of verbal inflection, as well as a reliance on copula verbs. This is illustrated in the following, a more or less continuous set of utterances by a single participant (ERI) at Time 2:

9. What the time now?
10. Who is er the birthday?
11. How many people the girl invited?
12. Who buy the cake?
13. Where is the equipment to play the music?
14. Why is the car broken?
15. ...why he or she is not er happy?
16. The house belong to who?
17. What are the words?

Evidence of omission or of double marking of the verbal element reflects patterns found in child L1 acquisition, and in the early stages of Pienemann's processability hierarchy at stages 1-3. This evidence suggested that there could be predictable evidence of a general linear progression (in line with Myles 2004 and Pienemann 1998) from chunking through omission, then oversuppliance to targetlikeness. The oral data from all participants were thus reanalysed comparing output at Time 1 and Time 2 to check for evidence of such progression.

The principal nontarget patterns related to lack of targetlike head movement or other issues related to verbal inflection: omitting tense marking or verbal element noted in Pienemann's hierarchy at Stage 1), leaving verb tensed but unraised or in-situ (Stage 2), oversupplying tense marking, e.g. using do-support as well as tense marking on verb (Stage 3). Other problems were switched use of *do* and *be,* and non-agreement of number (plural verb marking on singular subjects).

The table below summarises the oral output to show total utterances at Time 1 and Time 2 (including non-question forms), as well as the total of accurate question forms and non-target question forms. In comparison to accurate question forms, nontarget forms increased slightly from just under half the number of target forms to just over half by Time 2.

Table 25: Summary of oral output at Time 1 and Time 2

	total utterances	target forms	non-target forms (% of target forms)
Time 1	1452	367	178 (48.5%)
Time 2	1184	345	175 (50.7%)

Movement of the wh-item was also analysed for evidence of any lack of wh-fronting (following L1 transfer of wh-in situ), but wh-fronting was found to be virtually 100% systematic at both times (only two examples of wh-in situ were found, one using a possessive "whose" and one using "why"). In addition, there was ample evidence of repair for most participants, indicating a conscious awareness of some element of the target form, and a degree of success after monitoring in producing the target form. Totals for the six problem areas are detailed in the table below.

Table 26: Nontarget patterns of verbal inflection at Time 1 and Time 2

Total	Omission	In-situ	Over-suppliance	*Be* instead of *do*	*Do* instead of *be*	Number
Time 1 (/178)	39	30	31	45	14	19
Time 2 (/175)	61	38	18	34	13	11

These figures show, at Time 1, a heavy reliance on *be*-auxiliary usage as a default verbal element, and also that oversuppliance of verbal morphosyntax (double marking) or leaving the verbal element in-situ was nearly as common as omission.

By comparison, at Time 2, omission or in-situ morphosyntax was much greater than oversuppliance, and the increase in omission was paralleled by a reduction in oversuppliance. The reliance on default *be* decreased from Time 1, but was still evident.

For information, descriptive statistics for total nontarget forms and submeasures were obtained and are summarised in the tables below.

Table 27: Nontarget form scores at Time 1 and Time 2

	Mean	SD	Minimum	Maximum
Total nontarget forms Time 1	5.56	2.96	0	10
Total nontarget forms Time 2	5.47	3.59	0	15

Table 28: Nontarget form scores by submeasure at Time 1 and Time 2

	Mean	SD	Minimum	Maximum
Omission Time 1	2.29	1.40	1	5
Omission Time 2	2.65	2.19	1	9
In-situ Time 1	2.14	2.11	1	9
In-situ Time 2	2.24	1.30	1	6
Oversuppliance Time 1	1.55	0.89	1	4
Oversuppliance Time 2	1.20	0.56	0	2
Be instead of *do* Time 1	1.96	1.22	1	6
Be instead of *do* Time 2	1.89	0.68	1	3
Do instead of *be* Time 1	1.75	0.89	1	3
Do instead of *be* Time 2	1.18	0.40	1	2
Number Time 1	2.50	1.64	1	5
Number Time 2	1.38	0.52	1	2

Statistical analysis was run to compare the differences between Time 1 and Time 2 (with and without the outliers), but only the reduction in oversuppliance (from 31 tokens to 18) was significant, according to Wilcoxon signed rank tests ($p<.05$). This may have arisen from four extreme cases skewing the group data, as there were four participants at Time 1 who produced 2 or more tokens of oversuppliance, who decreased to 1 or 0 tokens by Time 2. I believe this change reflects a change in production strategy rather than an improvement from Stage 3 to Stage 4 (from double marking to targetlike copula questions). These four individuals still ranked well above mean for their overall nontarget total score at Time 2; three of the four also produced 2 or more tokens of in-situ verbal elements at Time 2.

In addition, at only 18 tokens out of 175, the use of oversuppliance across the group was much less common compared to other nontarget forms, and does not, I conclude, illustrate any robust trend of improvement.

The comparisons between Time 1 and Time 2 by type (using total data shown in Table 26 above) are illustrated for clarity in the figure below.[19]

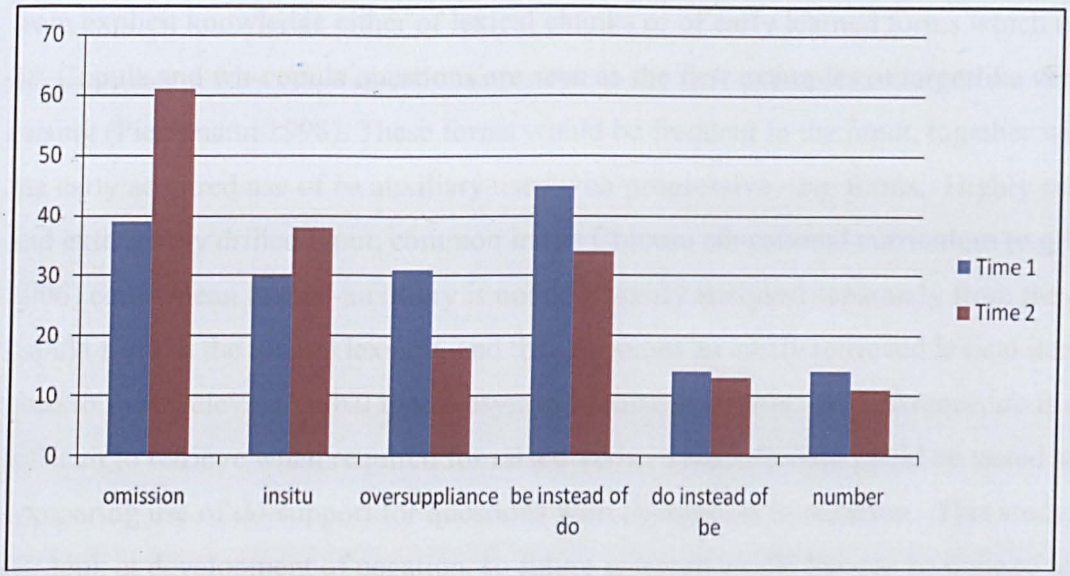


Figure 10: Comparison of nontarget patterns (Time 1 and Time 2)

These data suggest a range of issues confronting participants, in that wh-fronting caused no difficulty, but appropriate verb raising (head movement) was more problematic. Participants either omitted the verbal element, or, if they were aware that some kind of lexical entry was needed, appeared to show a preference for a default use of inflected *be*, or overcompensated by double-marking verbal inflections. By Time 2, in-situ tensed verbs were slightly preferred as an equivalent to default use of *be* (possibly as a result of one outlier's high use of in-situ forms), but omission of any verbal element increased. Evidence of oversuppliance decreased significantly ($p<.05$), again perhaps due to three outliers at Time 1 who decreased by Time 2.

A number of strategies are suggested to explain the patterns in the nontarget verbal inflection data shown here.[20]

---

[19] Raw totals aggregated across all participants' output are used here, rather than mean scores, to show the comparative range more clearly (since the mean scores all congregate between 1 and 2).

[20] It is noted that this analysis creates issues in distinguishing between acquisition of wh-movement per se, which incorporates the need for verb raising (at least for object questions), and acquisition of inflectional morphosyntax which can be argued to be a separate question from verb raising

There is some evidence of verb-raising to show verbal inflectional features of tense and number, although this is primarily shown through a default use of *be* in the raised position. I suggest that the reliance on *be* either as copula or as auxiliary is derived from explicit knowledge either of lexical chunks or of early learned forms which use *be*. Copula and wh-copula questions are seen as the first examples of targetlike verb raising (Pienemann 1998). These forms would be frequent in the input, together with the early acquired use of *be* auxiliary used with progressive *–ing* forms. Highly salient and extensively drilled input, common in the Chinese educational curriculum (e.g. Nani 2006) could mean that *be*-auxiliary is not necessarily analysed separately from the copula form in the mental lexicon, and thus becomes an easily retrieved lexical item used to show relevant verbal morphosyntax in raised position. By inference, *do* is more difficult to retrieve when required for raised verbs. This inference could be tested by comparing use of *do*-support for questions with *do*-support in negation. This study did not look at development of negation, so future research would have to be done to test this suggestion.

The emphasis on copula questions could have been assumed to reflect a L1-transfer effect of the use in Chinese of the existential verb *you* ("there is") in clause-initial or subject position (Yip 1995 – see discussion in chapter 2, section 2.2). If there is some kind of lexicalised L1/L2 form-meaning mapping between the use of existential "is there" (translating *you* directly) as a possible lexical question-marker, this could promote copula questions in the data, particularly if such questions occur with another lexical verb in canonical verb position. Some examples of such constructions were found (e.g. "Is there anybody bring chocolate with him?") but not significantly more using copula and lexical verbs in one token than using copula alone (one token was found among a random sample of ten participants). So a possible reliance on L1 transfer of existential *you* is not robustly sustained.

It is also possible that default *be* may be a generalisable feature of L2 English interlanguage regardless of the L1 and/or previous educational exposure. Evidence has been found in a number of studies(e.g. discussed in Hawkins 2001; see also Haznedar 2001; Ionin and Wexler, 2002; Hawkins and Casillas 2008) of early overgeneralisation of copula *be*. This could be taken as evidence in support of a structure-building approach, which could be seen in these studies as the earliest acquired marker of tense, acting as a trigger for the TP projection. This may be due to the copula being the "least specified" morpheme that can appear at TP and taking the widest range of complements (Hawkins 2001: 64).

However, there are a few difficulties with a structure-building approach. Firstly, the *be* morpheme was usually found to be correctly inflected despite other evidence of tense and agreement morphology being missing (in line with Haznedar 2001), arguing against it being a trigger for TP. In addition, there was wide prevalence of appropriate tensed verbs left in-situ, which is the opposite of what structure-building hypotheses would suggest. Further research teasing apart different contexts for tense marking and verb raising (such as acquisition of negation, mentioned above, or comparing declarative with interrogative contexts) would perhaps clarify whether use of default *be* and in-situ tensed verbs show some kind of implicational relationship.

Oversuppliance or omission of tense features were equally used as an alternative to in-situ tense marking at Time 1. In-situ and omission increased by Time 2, while oversuppliance decreased. Omission could be due to L1 transfer (as Mandarin does not mark Tense) but evidence of similar patterns in other L1 learners of English suggests that this is more akin to a L1-independent Basic Variety (Klein and Perdue 1997). This speculation is supported by studies of children with SLI that problems with tense marking, including on *be*-auxiliary is used as a marker of representational impairment (J. Paradis, 2007).

Oversuppliance is argued here to result from a combination of chunked knowledge of inflection in non-raised verbs, or feature-generated verb-raising, where both verbal elements can be easily processed at the same time, as found in Pienemann's Stage 3, which identifies oversuppliance as well as in-situ tensed verbs as interim strategies between omission and target-like raising, but Pienemann does not give any prediction or explanation for the optionality and default suppliance of *be* found in the data here. The decline in oversuppliance found in the data here could be explained by the increase in in-situ marking, as the four individual participants who showed most reliance on oversuppliance at Time 1 also showed increases in in-situ marking at Time 2.

Why some participants both omit and supply morphosyntax within the same dialogue is less clear. I infer, from the evidence here and from the hierarchy suggested in Processability theory, that omission is easier in processing terms.

The data suggests an implicational hierarchy of strategies to produce English questions which combine processing and acquisition constraints (similar to Pienemann's hierarchy but with more micro-analysis of why error and target forms occur in the same output). Wh-fronting appears first with omission of any related L2 morphosyntax element; it is easy to process and does not entail any acquisition; wh-fronting with verb-in situ is then acquired, but is processed less easily than omission; wh-movement which entails verb-raising features is then acquired but *be* is used as a default verb raising marker through greater ease of processing; finally wh-movement with relevant verb raising can be acquired through separation of *be* and *do* verbal markers.

Turning now to the question of oral fluency improvement, it is noted that the picture so far of lack of real improvement in total scores of accurate targetlike questions in the oral task was also reflected in the lack of change in accuracy in the RT task (see further discussion below), but the RT task did show significant change in faster speeds (as shown in the previous chapter). To see if this had any parallel in the oral data, a secondary set of analyses was carried out to investigate if performance in the oral output changed in any significant way, if measured in terms of overall fluency, especially using measures of hesitation and repair, or lexical diversity (Dechert 1980; Towell et al 1996; Malvern et al 2004). The oral data were transcribed according to standard conventions of oral analysis software based on the CHILDES project (CLAN), using two calculation programmes (frequency or FREQ, and Mean Length of Turn or MLT. These two programmes allowed for two measures to be used: type-token ratio, and hesitancy phenomena, comprising the number of repairs and filled pauses (adapted from Towell et al 1996).[21]

Type-token ratio was automatically calculated using the FREQ programme in CLAN. An example of how hesitancy phenomena were calculated is shown using the sample given below. Each pair of brackets () denotes a fragment. The FREQ programme

[21] The procedure for calculating the hesitancy measure was confirmed as suitable by A. David, p.c., conforming to standards of other fluency-type measures, used by oral corpora projects such as the French Language Learners' Oral Corpora project (www.flloc.soton.ac.uk)

calculates the total number of full words including fragments, and shows the distribution of filled pauses, specified in accordance with CLAN guidelines, to include *ah, aha, er, um, mm, mmhm, oh, uh*. The MLT programme calculates the total number of words without the fragments (i.e a lower number). Subtracting the lower MLT word total from the higher FREQ total provides the number of fragments. This number was added to the total number of filled pauses counted in the FREQ programme and yielded the total for repairs used here.

Example of CLAN-marked output (taken from a single participant's output at Time 2):

(1) **Er** <whe(**re**)> [/] <where is> [//] <where is her> [///] sorry **er** <how> [/] how much friend **er** of her will come?

(2) <Why I think> [//] **er** is that boy look very angry why?

(3) And what's the wor(d) <did she> [//] **er** was she prepared to say?

(4) <**Oh**> [/] <**oh**> [/] **oh** a student.

(5) But <if the> [//] <if> [/] if he <mi(**ss**)> [/] **er** miss the party <did> [/] <di(**d**)> [/] did they would separate?

Total hesitancy phenomena on this example: 9 filled pauses and 3 fragments = 12.

The data for mean scores on type-token ratio, and hesitancy phenomena are given in the table below, and were compared using Wilcoxon signed rank tests for significance.

Table 29: Analysis of oral fluency measures

	Mean	SD	Minimum	Maximum
Type token ratio Time 1	.409	.047	.333	.513
Type token ratio Time 2	.435	.050	.332	.563
Hesitancy phenomena Time 1	54.09	24.38	8	129
Hesitancy phenomena Time 2	42.38	15.32	10	78

Both measures showed significant improvement over time: the type-token ratio increased ($p<.001$), and the number of hesitancy phenomena decreased ($p<.05$). These data are illustrated below.

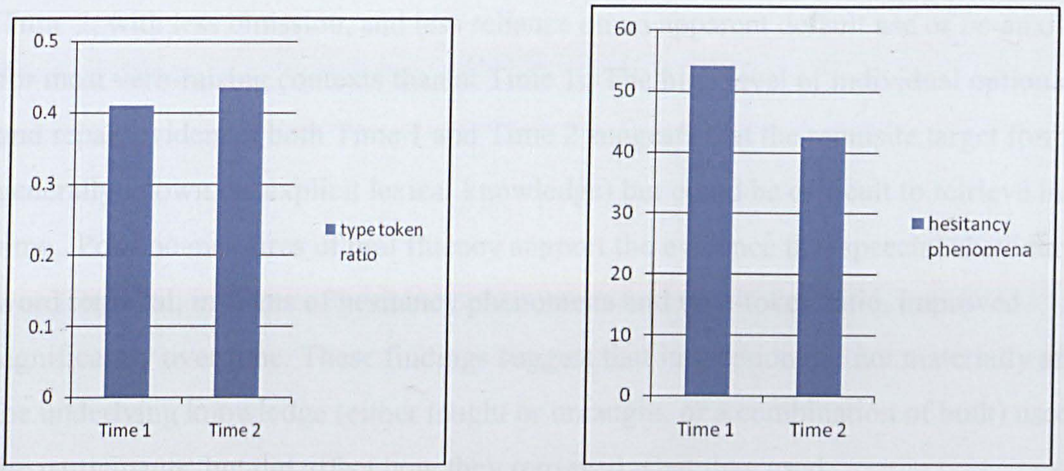


Figure 11: Comparison of changes in type/token ratio and hesitancy phenomena

In other words, the oral data showed significant improvement in measures that could be taken to denote greater fluency in oral performance, and may be taken to show some parallel with the findings in the RT task of faster speed, but no change in accuracy. However, it is noted that these measures only signify overall performance in the oral task, and may not be connected to the morphosyntactic questions under investigation, since the type-token ratio measure reflects growing general lexical diversity rather than morphosyntactic tokens per se.

Nevertheless, the hesitancy phenomena improvement shows, I believe, that speakers were using less hesitation, and less need to monitor their output, which could be taken as a hallmark of increasing linguistic proficiency or automaticity (Dechert 1980; Towell et al 1993; Herschensohn 1999; O'Brien et al 2006). It is important to note that the increase in omission of verbal marking may be connected with the decrease of hesitancy – speakers may simply be getting through their utterance with greater interest in pragmatic or communicative intent (Schauer 2004), than monitoring and repairing for inaccuracy of verbal marking.

To conclude, in the oral data, there was no clear pattern of development in accuracy between Time 1 and Time 2 across the whole group as predicted in the study's research hypotheses. However, posthoc analysis of non-target forms arguably showed evidence of an implicational hierarchy of processing strategies based on existing lexical knowledge, from omission through oversuppliance and optional suppliance to target-like production. The evidence here showed some progression through the hierarchy by

Time 2, with less omission, and less reliance on an apparent default use of *be*-auxiliary for most verb-raising contexts than at Time 1. The high level of individual optionality and repair evident at both Time 1 and Time 2 suggests that the requisite target form was generally known (as explicit lexical knowledge) but could be difficult to retrieve in real time. Posthoc measures of oral fluency support the evidence that speech behaviour and word retrieval, in terms of hesitancy phenomena and type-token ratio, improved significantly over time. These findings suggest that immersion did not materially affect the underlying knowledge (either taught or untaught, or a combination of both) used by the participants, but did affect how they retrieved what they used.

I turn now to analysis of the grammaticality judgement task to see how far these patterns were echoed in reaction time speed and accuracy.

5.3.2. Reaction Time Data

As outlined in section 5.2.2 above, this computerised grammaticality judgement task measured Reaction Times and targetlike accuracy on 68 tokens testing knowledge of wh-movement across three different wh-question structures: short movement (40 tokens), long movement (16 tokens), subjacency-constrained forms (12 tokens). The data were scored for total RT and accuracy on all 68 tokens, then subdivided for scores on grammatical compared to ungrammatical tokens (34 each), and short and long movement tokens were balanced for subject and object questions (28 each).

Hypothesis 1, relating to the RT task, was that participants would be more efficient in accessing existing knowledge of grammatical forms, especially taught short movement question forms, than either complex long movement questions or untaught implicit ungrammatical forms, including subjacency violations, and in processing object questions compared to subject questions, measured in lower RTs and higher scores of targetlike accuracy in grammaticality judgements. Hypothesis 2, relating to the RT task, was that participants would improve in their knowledge and use of question forms (subject to the asymmetries noted in Hypothesis 1) after exposure to enriched input in an immersion setting.

The data are presented below for Time 1 and Time 2, showing scores for RT and accuracy measures, together with scores for grammatical and ungrammatical tokens.

The results for the three question types and object and subject sub-divisions are presented separately. Data relating to RT speeds are presented first, followed by the data for accuracy.

5.3.2.1. RT speeds

All RT data were scored by the computer clock in milliseconds so that results are calculated in milliseconds, but data are reported here in seconds for ease of reference.[22] The tables below show statistical data for Time 1 and Time 2; means are compared in Figure 4 below to illustrate the differences of RTs by type, and in comparison between Time 1 and Time 2. Note that improvements in RT show as lower times (i.e. faster judgements)[23].

Table 30: RT (secs) for total, grammatical and ungrammatical tokens at Time 1

	Mean	SD	Minimum	Maximum
RT Total	506.37	159.17	155.04	846.19
RT grammatical	248.24	83.58	77.34	458.11
RT ungrammatical	258.13	78.80	77.70	415.28

Table 31: RT for total, grammatical and ungrammatical tokens at Time 2

	Mean	SD	Minimum	Maximum
RT Total Time 2	432.81	151.75	235.90	997.02
RT gram Time 2	208.14	73.20	113.42	436.99
RT ungram Time 2	224.68	81.38	117.81	560.03

Table 32: RT by submeasure at Time 1

	Mean	SD	Minimum	Maximum
RT Short movement	239.74	68.77	79.10	367.20
RT Long movement	115.55	49.61	32.49	277.37
RT Subjacency	93.22	35.04	24.86	188.56
RT Object questions	174.89	57.66	57.61	337.79
RT Subject questions	181.70	60.06	53.97	306.78

22 All the data analyses are in ms, avoiding any rounding up errors distorting the calculations.

23 Exploratory analysis revealed two outliers, one at each extreme of the RT scores, who scored just more than 2SD beyond the mean; however, excluding them did not affect the patterns of significance found in the scores for Time 2 or in the change in scores, so they are retained in the analysis shown here.

Table 33: RT by submeasure at Time 2

	Mean	SD	Minimum	Maximum
RT Short movement Time 2	204.00	67.22	115.02	428.95
RT Long movement Time 2	94.91	35.58	47.05	219.56
RT Subjacency Time 2	81.10	34.37	40.79	230.39
RT Object questions Time 2	143.56	43.55	84.65	276.63
RT Subject questions Time 2	155.34	58.92	79.46	371.88

The data show a marked improvement in RT with lower mean RT scores by Time 2 found for all measures, illustrated in Figure 6 below.

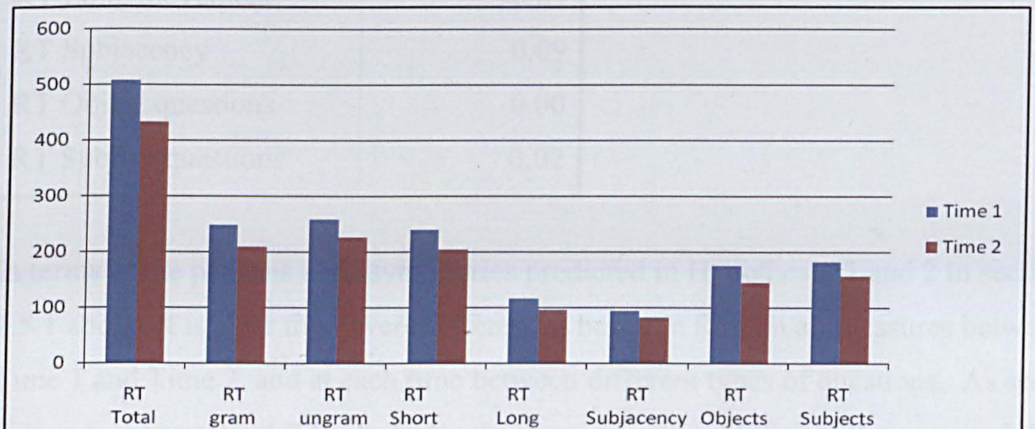


Figure 12: Mean RT scores (in seconds) at Time 1 and Time 2

The changes in scores over time are presented for all measures in the table below.

Table 34: Change in scores between Time 1 and Time 2

	Mean	SD	Minimum	Maximum
Change RT total	-73.56	183.58	-368.07	435.30
Change grammatical RT	-40.10	92.83	-203.53	214.77
Change ungrammatical RT	-33.46	93.40	-175.88	220.53
Change Short movement RT	-35.74	76.31	-154.45	202.23
Change Long movement RT	-20.64	51.10	-128.28	92.40
Change Subjacency RT	-12.12	42.70	-94.72	108.43
Change Object RT	-31.33	57.51	-138.75	137.99
Change Subject RT	-26.36	66.35	-129.67	157.24

RT speeds at Time 1 and Time 2 were compared for significant differences using Wilcoxon signed rank analysis, and all improvements in speed were highly significant, apart from Subjacency, as shown in the table below.

Table 35: Analysis of significant difference in RT scores, Time 1 and Time 2

	Sig. (2-tailed)
RT Total	0.01
RT grammatical	0.01
RT ungrammatical	0.01
RT Short movement	0.00
RT Long movement	0.03
RT Subjacency	0.09
RT Object questions	0.00
RT Subject questions	0.02

In terms of the patterns and asymmetries predicted in Hypotheses 1 and 2 in section 2.5.1 above, it is clear there were differences between RTs on all measures between Time 1 and Time 2, and at each time between different types of questions. As seen in Figure 6, grammatical RT was faster than ungrammatical RT at both times, and this difference was significant ($p<.05$ at Time 1, $p<.01$ at Time 2) according to Wilcoxon signed rank test analysis. In terms of change over time, as seen in Table 34 above, grammatical RT improved more than ungrammatical RT (-40.10 seconds for grammatical RT, compared to -33.46 for ungrammatical RT), although this was not statistically significant ($p>.1$). Object questions were also judged faster than subject questions at both times; this difference was not significant at Time 1 ($p>.1$) but was significant at Time 2 ($p<.01$). This seems to be due to a greater improvement in Object RT (-31.33 seconds) than Subject RT (-26.36 seconds), although, again, the greater change in Object RT was not significant ($p>.4$).

In order to compare the three specific sub-measures, which were not balanced for number of tokens, RTs for each submeasure were recalculated giving an average RT by item, to assess the average speed by submeasure at Time 1, Time 2 and change in RT over the period of immersion, shown in the table and figure below.

Table 36: Average speed (in secs) by submeasure at Time 1, Time 2, and change

	Time 1	Time 2	Change
Average RT Short movt	5.993	5.100	-0.894
Average RT Long movt	7.222	5.932	-1.290
Average RT Subjacency	7.768	6.758	-1.010

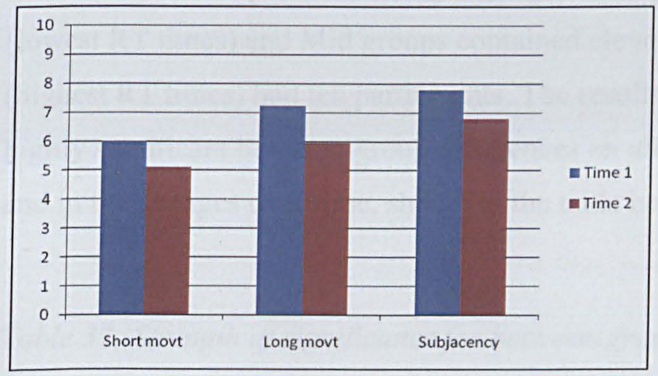


Figure 13: Average item speed (in seconds) by submeasure

This confirmed that short movement, descriptively, was judged quicker than all other types at Time 1 and Time 2, but that long movement improved the most by Time 2. The RT data were also converted to z-scores, and compared using Wilcoxon signed rank analysis, but there were no statistically significant differences between each submeasure ($p>.5$) at either Time 1 or Time 2 or in the change in RT over time.

The findings described above provide evidence to support Hypothesis 1 in that participants across the whole group showed predictable patterns of asymmetry: at both Time 1 and Time 2 short movement items (i.e. simple questions) were judged faster than long movement items (i.e. complex questions) and subjacency-constrained items, although these asymmetries were not always statistically significant. In contrast, grammatical items were judged significantly faster than ungrammatical items at both times ($p<.05$). The predicted object/subject asymmetry was also found, and was statistically significant at Time 2 ($p<.01$). Hypothesis 2 was also supported with statistically significant improvement shown in overall speed by Time 2 ($p<.01$), and across nearly all sub-measures, and followed the pattern of asymmetries shown above. The greatest improvement was found in faster judgements on grammatical long movement items, shown in significant improvements for both long movement ($p<.05$) and grammatical items ($p<.01$), seen in Table 35.

Turning now to between-group analysis to compare the overall patterns discussed above, this analysis was done to see if the lack of significance found between the three syntactic submeasures of short movement, long movement and subjacency for the whole pool of participants overall obscured significant individual variation. Participants were split into three groupings, by level of RT speed at Time 1. The High (lowest RT times) and Mid groups contained eleven participants, and the Low group (highest RT times) had ten participants. The results of a Kruskal-Wallis test showed highly significant between-group differences on all measures at both Time 1, Time 2 and in the changes over time, shown in the table below.

Table 37: Strength of significance for between-group differences on RT submeasures at Time 1, Time 2 and change

	Sig. (shown in bold if $p<.05$)		
Submeasure of Accuracy	Time 1	Time 2	Change
RT Short movement	**0.00**	**0.03**	**0.03**
RT Long movement	**0.00**	0.06	**0.02**
RT Subjacency	**0.00**	0.10	0.06

To identify which groups were showing significant differences, an analysis of mean scores was run for each submeasure at Time 1 and Time 2 and for change by Time 2, split into the same three groups in the Kruskal-Wallis test. These data are shown in raw scores (not z-scores[24]) in the tables below.

[24] Raw scores are used rather than z-scores for greater exactitude of comparison for each submeasure between Time 1 and Time 2, as the data here are to show between-group differences on each measure, not between-measure differences, due to the greater number of short movement tokens.

Table 38: Mean scores by group for each submeasure at Time 1

Time 1		Mean	SD	Minimum	Maximum
Low	RT Short movement	309.51	35.35	256.35	367.20
	RT Long movement	167.50	47.19	127.16	277.37
	RT Subjacency	131.47	28.01	86.27	188.56
Mid	RT Short movement	250.32	28.13	195.91	288.85
	RT Long movement	114.00	16.67	88.05	150.56
	RT Subjacency	92.05	12.35	75.94	113.18
High	RT Short movement	165.73	41.58	79.10	219.63
	RT Long movement	69.86	18.71	32.49	91.57
	RT Subjacency	59.62	15.49	24.86	75.10

Table 39: Mean scores by group for each submeasure at Time 2

Time 2		Mean	SD	Minimum	Maximum
Low	RT Short movement Time 2	235.62	84.23	163.42	428.95
	RT Long movement Time 2	113.04	49.15	53.11	219.56
	RT Subjacency Time 2	98.18	52.10	40.79	230.39
Mid	RT Short movement Time 2	213.04	49.71	155.41	317.34
	RT Long movement Time 2	96.40	25.81	55.77	141.55
	RT Subjacency Time 2	80.82	20.26	52.57	117.04
High	RT Short movement Time 2	166.21	49.91	115.02	281.33
	RT Long movement Time 2	76.94	19.94	47.05	114.48
	RT Subjacency Time 2	65.85	16.12	42.00	93.90

Table 40: Change in RT scores by group for each submeasure

		Mean	SD	Minimum	Maximum
Low	Change Short movement RT	-73.88	88.71	-154.45	148.69
	Change Long movement RT	-54.46	65.20	-128.28	92.40
	Change Subjacency RT	-33.29	62.19	-94.72	108.43
Mid	Change Short movement RT	-37.28	46.39	-100.99	28.49
	Change Long movement RT	-17.60	36.02	-94.80	28.21
	Change Subjacency RT	-11.23	23.00	-50.72	28.06
High	Change Short movement RT	0.48	76.92	-79.19	202.23
	Change Long movement RT	7.08	31.17	-32.06	82.00
	Change Subjacency RT	6.23	28.70	-32.55	69.03

These between-group comparisons shown in the tables above are also illustrated in the figure below.

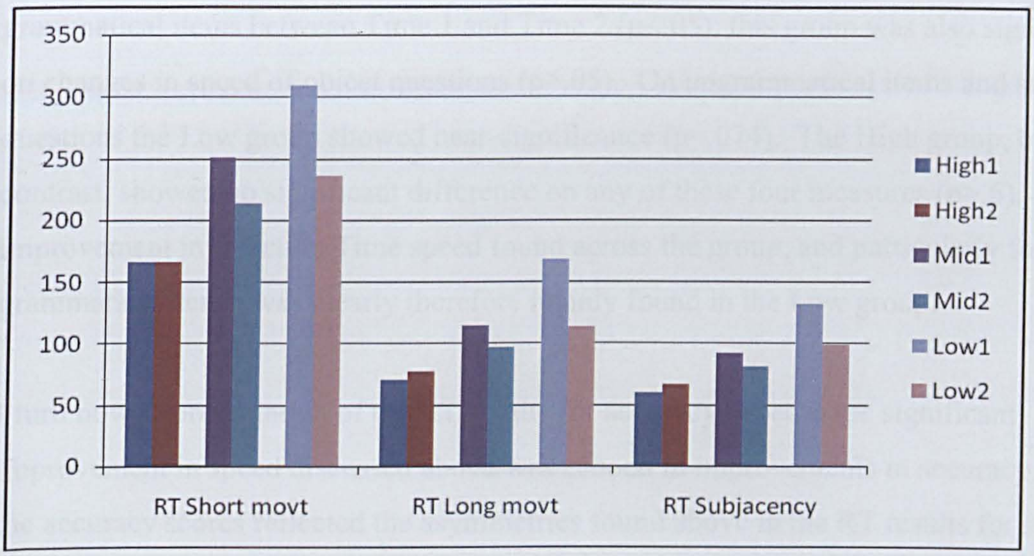


Figure 14: Mean RT (in secs) by group across submeasure, Time 1 and Time 2

The analysis shown in the tables and figures above illustrates the significant between-group differences shown in the Kruskal-Wallis test. The High group (lowest RT times) remained the fastest across all measures at both Time 1 and Time 2. However, in terms of change, a number of different patterns of between-group variation were found, with the Low group (with the highest RT times) showing most change on each submeasure, and the High group showing either no improvement or a slight increase in RT time on each submeasure. Wilcoxon signed rank analysis of difference in RT between Time 1 and Time 2, split by group, confirmed that for the High group, there were no significant differences on any submeasure ($p>.4$). The Mid group improved significantly only on short movement ($p<.05$), and for the Low group the change on short movement and long movement approached significance ($p=.059$), but the change in subjacency was not significant ($p>.1$).

To sum up the analyses of the RT speed data, the patterns analysed above confirmed the pattern of asymmetry found across the whole group, but indicated that the greatest improvement in speed was found among those who were slowest at Time 1, with the High group showing no evidence of improving across the different types. Between-group differences were analysed to also further assess the evidence shown earlier that the greatest improvement for the whole group was found on grammatical items. According to Wilcoxon signed rank analysis on grammatical items, ungrammatical items and subject and object questions, the Low group showed significant difference on grammatical items between Time 1 and Time 2 ($p<.05$); this group was also significant on changes in speed of object questions ($p>.05$). On ungrammatical items and subject questions the Low group showed near-significance ($p=.074$). The High group, by contrast, showed no significant difference on any of these four measures ($p>.6$). The improvement in Reaction Time speed found across the group, and particularly for grammatical items, was clearly therefore mainly found in the Low group.

I turn now to presentation of the RT results for accuracy to see if the significant improvement in speed discussed above was echoed in improvements in accuracy, and if the accuracy scores reflected the asymmetries found above in the RT results for speed.

5.3.2.2. RT accuracy

These data are presented in the tables below in similar form to the results presented for the RT speed data, with group analysis for Time 1 and Time 2 of the overall total accuracy scores, together with the scores for grammatical and ungrammatical items (Tables 41 and 42). The results for the three question submeasures and object and subject sub-divisions are presented separately (in Tables 43 and 44). As outlined in section 2.2, the total number of tokens being tested were 68, balanced for grammaticality. There were 40 Short movement tokens, 16 Long movement tokens, 12 Subjacency-constrained tokens and 28 Object and Subject questions.

Table 41: Accuracy for total, grammatical and ungrammatical tokens at Time 1

	Mean	SD	Minimum	Maximum
Accuracy (/68)	40.13	10.76	20	57
Grammatical (/34)	22.22	7.86	4	38
Ungrammatical (/34)	17.91	8.27	0	32

Table 42: Accuracy for total, grammatical and ungrammatical tokens at Time 2

	Mean	SD	Minimum	Maximum
Accuracy 2	39.69	8.626	18	56
Grammatical 2	23.59	6.603	9	36
Ungrammatical 2	16.09	6.177	5	33

Table 43: Accuracy by submeasure at Time 1 (/total number of tokens per submeasure)

	Mean	SD	Minimum	Maximum
Short movement (/40)	22.97	6.54	10	36
Long movement (/16)	7.50	3.77	1	16
Subjacency (/12)	5.38	3.16	0	12
Object questions (/28)	16.94	4.50	4	24
Subject questions (/28)	13.47	5.04	4	22

Table 44: Accuracy by submeasure at Time 2

	Mean	SD	Minimum	Maximum
Short movement 2	23.47	4.88	12	34
Long movement 2	8.19	3.35	2	15
Subjacency 2	4.31	2.57	1	10
Object questions 2	16.81	3.62	10	24
Subject questions 2	14.84	4.10	5	23

These data show a less clear picture than the RT speed data, with a slight decrease in total accuracy, derived from decreased accuracy on ungrammatical items, primarily on subjacency, together with a very slight decrease on object accuracy. However, the mean overall scores were above chance at Time 1 and Time 2 (although just below chance for ungrammatical items at Time 2). There was also a notable increase in minimum scores, all of which increased in accuracy even for subjacency and object questions. By Time 2, only eight participants were scoring below chance, meaning that two-thirds of all participants (twenty four) were scoring above chance. Fourteen scored 41 or above (i.e. 60% accuracy).

The data for Time 1 and Time 2 are also illustrated for clarity of comparison in the figure below.

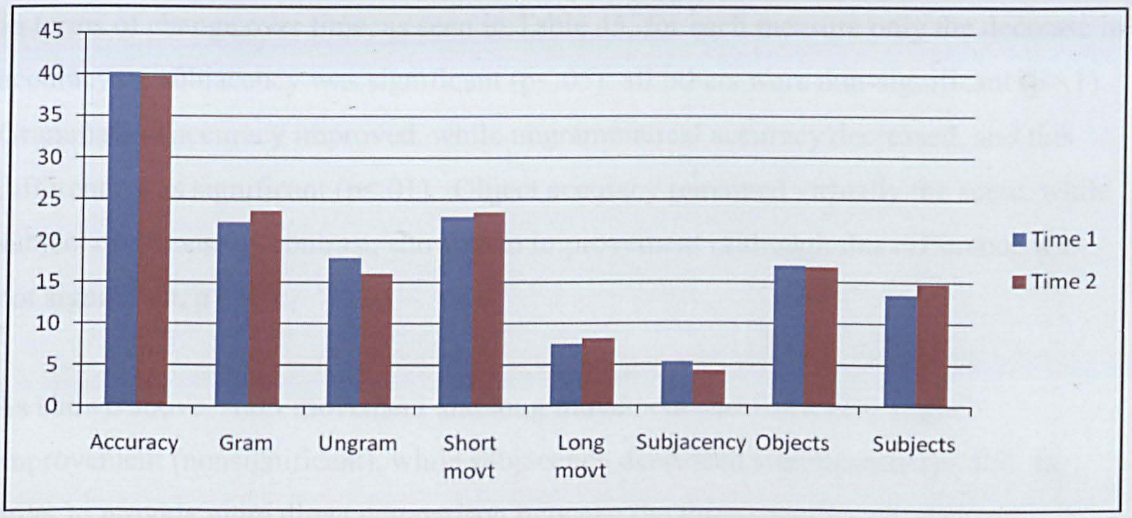


Figure 15: Mean scores in RT Accuracy at Time 1 and Time 2

The scores for change in accuracy scores across the submeasures by Time 2 are shown in the table below.

Table 45: Change in Accuracy scores by Time 2

	Mean	SD	Minimum	Maximum
Change Accuracy	-0.44	12.16	-29	23
Change Grammatical accuracy	1.38	7.29	-16	16
Change Ungrammatical accuracy	-1.81	8.24	-15	14
Change Short movement accuracy	0.50	7.72	-19	15
Change Long movement accuracy	0.88	3.20	-10	8
Change Subjacency accuracy	-0.75	3.07	-5	5
Change Object accuracy	-0.13	5.14	-10	12
Change Subject accuracy	1.38	6.08	-12	17

In terms of the patterns and asymmetries predicted in Hypotheses 1 and 2 above, comparing Time 1 scores and Time 2 scores, there were differences at both times between the various submeasures. As shown in the preceding tables and figure, grammatical items were more accurate at both times than ungrammatical items, and object questions were more accurate than subject questions, and these differences were highly significant, according to Wilcoxon signed rank tests (p<.01).

In terms of change over time, as seen in Table 45, for each measure only the decrease in accuracy on subjacency was significant (p<.05); all others were non-significant (p>.1). Grammatical accuracy improved, while ungrammatical accuracy decreased, and this difference was significant (p<.01). Object accuracy remained virtually the same, while subject questions, by contrast, showed an improvement (although this difference was not significant, p>.05).

As shown above, short movement and long movement showed a very slight improvement (nonsignificant), while subjacency decreased significantly (p<.05). In order to provide more direct comparison between the three specific sub-measures, which were not balanced for number of tokens, accuracy mean scores for each submeasure were converted into percentages by dividing the raw scores by the total for each submeasure, yielding comparable data, shown in the table and figure below.

Table 46: Comparison in percentage terms across submeasures

Accuracy (%)	Time 1	Time 2	Change
Short movement	57.42%	58.67%	1.25%
Long movement	46.88%	51.17%	5.47%
Subjacency	44.79%	35.94%	-6.25%

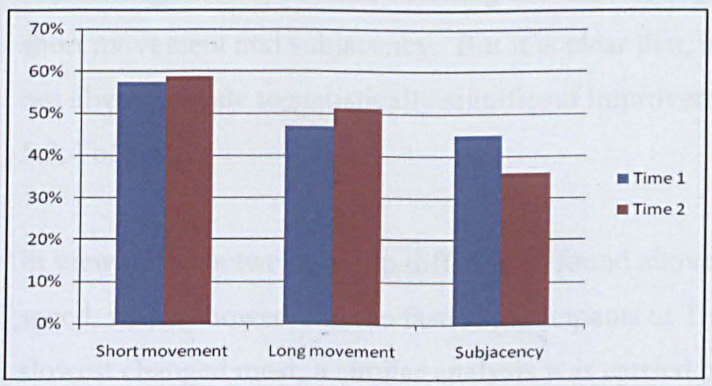


Figure 16: Comparison of accuracy in percentage terms across submeasures

The data discussed here provide further evidence of the asymmetries argued in the research hypotheses, that simple forms (short movement) would be more accurate than complex forms, and that both forms would be more accurate than nontaught forms (subjacency-constrained) at both times. The data presented in Table 46 also show that long movement improved the most (over 5%), and subjacency decreased the most (-6.25%). The data were then converted to z-scores, and checked for significant differences between each type, but the descriptive asymmetries shown above were nonsignificant (p>.5).

To sum up the overall findings for accuracy, the data largely echo the results of RT speed, discussed in section 5.3.2.1 above, and support the asymmetries predicted in Hypothesis 1. Short movement items were judged faster and more accurately than long movement items and subjacency-constrained items, although this asymmetry was non-significant (p>.05). Grammatical items were significantly both faster and more accurate than ungrammatical items (p<.05). Objects were also judged faster and more accurately than subjects (significantly at both times for accuracy and significantly at Time 2 for speed, p<.01).

However, the patterns of change over time in accuracy were not as clear as for speed. Overall there was a slight decrease in accuracy, particularly in ungrammatical items, which could be reflected in the smaller improvement in RT speed for ungrammatical items compared to grammatical items (shown in Table 34 above). The significantly faster speeds found on grammatical items and long movement were echoed in significantly more accurate responses on grammatical items, and in relatively greater (but nonsignificant) accuracy on long movement, in percentage terms, compared to short movement and subjacency. But it is clear that, in group terms, faster speeds did not always equate to statistically significant improvements in accuracy (see section 5.3.4 below).

In view of the between-group differences found above in section 5.3.2.1 on changes in speed, which showed that the fastest participants at Time 1 changed least, and the slowest changed most, a similar analysis was carried out on the accuracy data. Participants were divided into three groups by accuracy scores at Time 1, split by equal percentiles. In this way the overall apparent lack of significant changes could be assessed for between-group significant variation in change, similar to the RT speed data. The Low group contained twelve participants, the Mid group had thirteen and the

High group had seven participants. There were a number of significant between-group differences, according to a Kruskal-Wallis test, as seen in the table below.

Table 47: Strength of significance for between-group differences in accuracy measures at Time 1, Time 2 and change over time

	Sig. (shown in bold if p<.05)		
Submeasure of Accuracy	Time 1	Time 2	Change
Grammatical questions	0.07	0.29	0.21
Ungrammatical questions	**0.00**	0.59	**0.00**
Short movement	**0.00**	0.26	**0.00**
Long movement	**0.04**	0.20	0.20
Subjacency	**0.00**	0.44	**0.02**
Object questions	**0.00**	0.50	**0.03**
Subject questions	**0.00**	0.28	**0.03**

All measures were significantly different at Time 1 apart from grammatical questions, but none were different at Time 2. In other words, the overall patterns in accuracy at Time 2 shown above did not obscure between-group differences at Time 2.

There were between-group differences in how the groups changed on at least some measures. There were no significant differences between groups on change in grammatical questions and change in long movement questions. There were, however, significant between-group differences for change in ungrammatical questions, short movement, subjacency, object questions and subject questions.

These statistical data are shown in the tables below.

Table 48: Between-group differences in accuracy by submeasures at Time 1

Time 1		Mean	SD	Minimum	Maximum
Low	Grammatical questions	18.25	9.99	4	38
	Ungrammatical questions	10.33	5.28	0	16
	Short movement	16.33	4.36	10	23
	Long movement	5.92	4.60	1	16
	Subjacency	3.33	1.97	0	6
	Object questions	12.75	4.29	4	18
	Subject questions	9.33	4.72	4	18
Mid	Grammatical questions	24.08	5.50	14	34
	Ungrammatical questions	20.08	5.77	11	29
	Short movement	25.38	2.84	22	31
	Long movement	8.77	2.98	4	15
	Subjacency	5.54	3.07	1	10
	Object questions	19.08	2.18	16	22
	Subject questions	15.08	2.96	11	21
High	Grammatical questions	25.57	4.86	17	33
	Ungrammatical questions	26.86	4.14	20	32
	Short movement	29.86	3.72	27	36
	Long movement	7.86	2.85	5	13
	Subjacency	8.57	2.30	6	12
	Object questions	20.14	2.12	18	24
	Subject questions	17.57	3.69	12	22

Table 49: Between-group differences in accuracy by submeasures at Time 2

Time 2		Mean	SD	Minimum	Maximum
Low	Grammatical questions Time 2	21.50	6.91	10	32
	Ungrammatical questions Time 2	15.67	3.87	10	21
	Short movement Time 2	22.33	4.40	14	31
	Long movement Time 2	7.33	3.55	2	14
	Subjacency Time 2	4.25	1.86	2	8
	Object questions Time 2	15.92	3.50	11	22
	Subject questions Time 2	13.75	4.07	7	22
Mid	Grammatical questions Time 2	25.38	7.17	9	36
	Ungrammatical questions Time 2	14.85	6.14	5	25
	Short movement Time 2	23.54	5.97	12	34
	Long movement Time 2	9.38	3.20	3	15
	Subjacency Time 2	3.69	2.66	1	9
	Object questions Time 2	17.08	4.33	10	24
	Subject questions Time 2	15.85	4.78	5	23
High	Grammatical questions Time 2	23.86	4.38	19	30
	Ungrammatical questions Time 2	19.14	8.93	9	33
	Short movement Time 2	25.29	3.15	21	30
	Long movement Time 2	7.43	2.99	4	12
	Subjacency Time 2	5.57	3.31	2	10
	Object questions Time 2	17.86	2.19	15	22
	Subject questions Time 2	14.86	2.48	13	19

Table 50: Between-group differences for change in accuracy by submeasure

Change		Mean	SD	Minimum	Maximum
Low	Change grammatical questions	3.25	8.83	-14	16
	Change ungrammatical questions	5.33	5.57	-5	14
	Change Short movement	6.00	5.74	-2	15
	Change Long movement	1.42	4.32	-10	8
	Change Subjacency	0.92	2.91	-4	5
	Change Object questions	3.17	4.75	-4	12
	Change Subject questions	4.42	5.28	-5	17
Mid	Change grammatical questions	1.31	6.86	-16	10
	Change ungrammatical questions	-5.23	6.42	-15	3
	Change Short movement	-1.85	8.15	-19	12
	Change Long movement	1.08	2.29	-2	5
	Change Subjacency	-1.08	3.01	-5	5
	Change Object questions	-2.00	4.90	-10	8
	Change Subject questions	0.77	6.22	-12	11
High	Change grammatical questions	-1.71	4.39	-7	4
	Change ungrammatical questions	-7.71	6.58	-15	2
	Change Short movement	-4.57	3.46	-10	1
	Change Long movement	-0.43	2.37	-4	3
	Change Subjacency	-3.00	1.83	-5	0
	Change Object questions	-2.29	3.50	-7	4
	Change Subject questions	-2.71	4.89	-8	5

As these tables showed, the only significant between-group differences were found for change in ungrammatical questions, short movement, subjacency, object questions and subject questions. These data are illustrated in the figures below, in which the three syntactic submeasures of short movement, long movement and subjacency are shown separately for clarity.

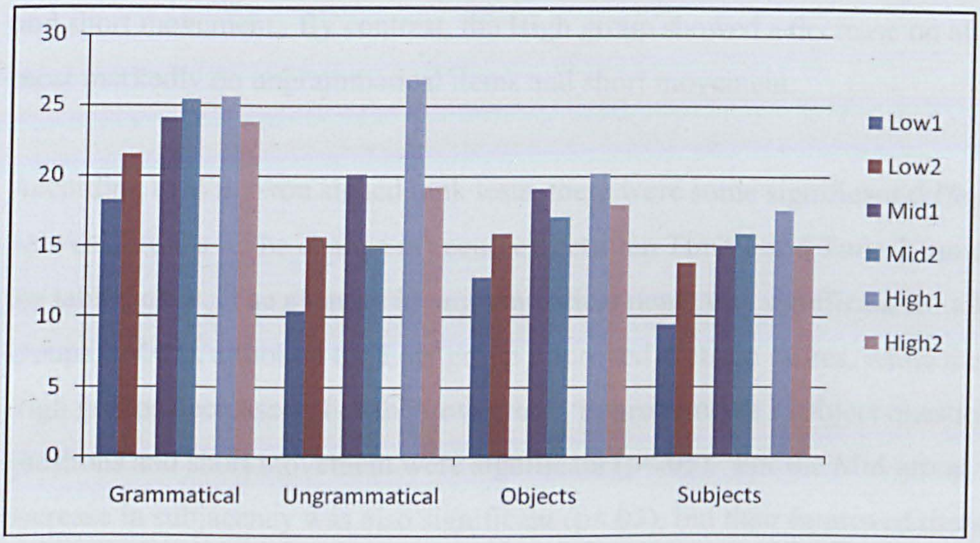


Figure 17: Mean accuracy scores by group for grammatical/ungrammatical and object/subject questions comparing Time 1 and Time 2

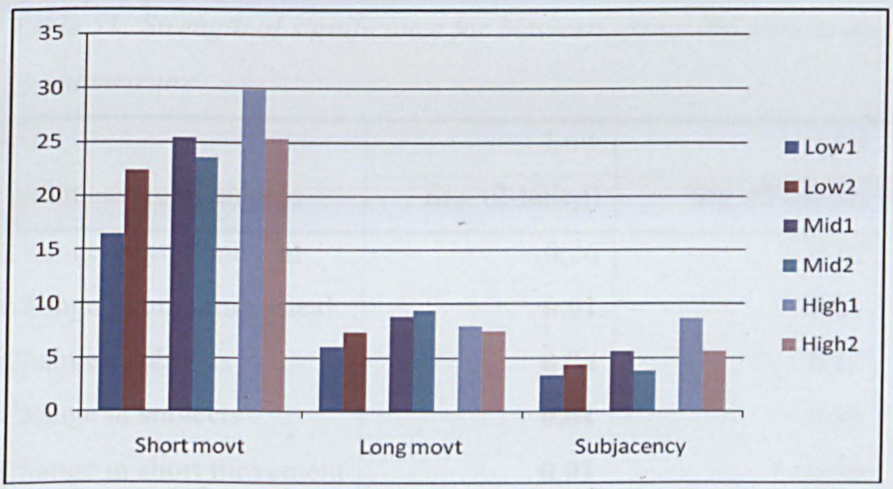


Figure 18: Mean accuracy scores by group for short movement, long movement and subjacency comparing Time 1 and Time 2

The tables above (Tables 48-50) providing the statistical data comparing the different groups, show that the groups behaved markedly differently across the submeasures, especially in how they changed in accuracy over time. The Kruskal-Wallis test had shown no between-group significant difference on change in grammatical and long movement items, although descriptively the Low and Mid groups both improved, while the High group decreased. All other measures of change had shown between-group significance ($p<.05$), and again this is found in the improvements on all measures shown by the Low group, most clearly on ungrammatical questions, subject questions and short movement. By contrast, the High group showed a decrease on all measures, most markedly on ungrammatical items and short movement.

According to Wilcoxon signed rank tests, there were some significant differences found between groups in the change in accuracy between Time 1 and Time 2, summed up in the table below. The changes in ungrammatical items was significant for all three groups ($p<.05$), although the Low group improved in mean scores, while the Mid and High groups decreased. For the Low group, improvements in object questions, subject questions and short movement were significant ($p<.05$). For the Mid group, the decrease in subjacency was also significant ($p<.05$), but their improved measures in grammatical items, subjects and long movement were non-significant. For the High group, the decreases in short movement and subjacency were significant ($p\leq.05$).

Table 51: Strength of significance for between-group differences on change in accuracy by submeasure

	Low	Mid	High
Submeasure of change	Sig. (2-tailed)	Sig. (2-tailed)	Sig. (2-tailed)
Change in grammatical	0.26	0.25	0.40
Change in ungrammatical	**0.01**	**0.01**	**0.05**
Change in objects	**0.04**	0.12	0.17
Change in subjects	**0.01**	0.59	0.18
Change in short movement	**0.01**	0.48	**0.03**
Change in long movement	0.09	0.39	0.60
Change in subjacency	0.28	**0.04**	**0.03**

To recap the overall and between-group analyses across the different submeasures of RT Accuracy, mixed evidence was found in line with the asymmetries expected in Hypothesis 1 and 2. As expected, short movement (i.e. explicit taught simple forms) was more accurate at both Time 1 and Time 2 across the whole group (although not significantly); however, the groups behaved differently over time, in that the Low group improved significantly while the High group decreased significantly. Long movement (i.e. complex questions) showed significant improvement across all groups by Time 2. By contrast, subjacency-constrained questions (i.e. implicit untaught forms) showed a decrease in accuracy overall, which was significant for the Mid and High groups, although the Low group showed some non-significant improvement. Object questions were significantly more accurate than subjects across all groups at both Time 1 and Time 2; although there was a small decrease in object questions across the whole group by Time 2, and a non-significant improvement in subject questions, the Low group showed significant improvement in both object and subject questions.

5.3.3. Conclusion of linguistic data analysis

Reviewing all the data discussed in the previous sections, there appears to be clear support for Hypothesis 1 (although not always statistically significant), but the evidence is mixed for the expected improvement over the year-long period of immersion predicted in Hypothesis 2, arising from significant between-group variation in how different groups improved or decreased over time. The robust significant improvement found in RT speeds, which could be read as evidence of acquisition or development of greater automaticity, does not appear to be matched by the patterns shown in the RT accuracy data.

It was plausible to suppose that there may have been a trade- off between accuracy and speed: in other words, that faster times came at the expense of accuracy, or, perhaps, that greater accuracy require slower speeds. However, Spearman tests run between RT speed and accuracy at both Time 1 and Time 2, and between change in Accuracy and change in Speed showed no significant correlations (p>.1). Therefore it does not seem as though participants were trading off speed and accuracy.

These overall conclusions from the RT task showed some similarity with the lack of significant improvement found over time in the oral data, which found some slight improvement in question/utterance ratio scores, but no significant improvement on any of the specific submeasures analysed.

The main conclusion to draw from the linguistic data overall is that immersion did not seem to aid acquisition overall (in terms of triggering more target-like oral production and RT judgements), but seemed to facilitate better use of existing knowledge, in terms of faster RT speeds, especially on grammatical items, and slightly higher question/utterance ratio and fewer nontarget forms in the oral task. Only the Low groups showed some significant improvements in both linguistic tasks, suggesting that two-thirds of the participants may have hit some kind of plateau even prior to immersion, and the amount of exposure gained during immersion was insufficient to change their underlying linguistic knowledge. This will be discussed in more detail in the next chapter.

However, the variation in improvements and decrease in linguistic scores on some measures could be expected to affect the correlations with Working Memory (WM) which formed the heart of this study. I turn now to analysis of the WM data in order to investigate how far the assumption of some implication for WM in linguistic development was supported by the data.

5.4. WM correlations

The role of WM in L2 development formed the basis of Hypotheses 3 and 4 of this study, repeated here:

3. WM capacity is implicated in participants' ability to access existing knowledge of taught question forms more efficiently, in that greater WM capacity would correlate with individual differences in rates of improvement in perception/production of targetlike question forms (apart from subjacency-constrained items) in both oral and timed grammaticality judgement tasks.
4. WM capacity is not implicated in their capacity to acquire untaught implicit subjacency constraints in a timed grammaticality judgement task.

In terms of answering these hypotheses about the implication of WM on linguistic development, statistical analysis of the WM data was run, first to show the descriptive baseline results for mean scores, SD and range at Time 1 and Time 2, and then to test for correlations between the linguistic scores and the WM scores. The following section 4.1 provides the baseline results at Time 1, Time 2 and any changes over time, and then section 4.2 provides correlational analysis between WM and the linguistic measures using Spearman non-parametric tests, first with the oral data, then with RT speed and accuracy.

As set out in section 2.3 at the start of this chapter, there were five WM tasks: the Listening Span task (in L2 English), Digits Back in L1 (Chinese) and L2 (English), Story Recall in L1 and L2. The Listening Span task and the Digits Back tasks were scored following Conway et al. (2005) as partial-credit scoring (taking all answers into consideration, as a proportion of the total number of items recalled), producing a ratio between 0 to 1 (see chapter 3 for scoring procedure in full). The Story Recall tasks were scored by accuracy of recall of both gist and grammatical phrasing out of a possible maximum of 50, but also shown here as ratio scores to provide comparable cross-reference with the other tasks.

5.4.1. WM mean scores at Time 1 and Time 2

Results for mean scores on all the WM tasks at Time 1 and Time 2 are shown in the tables below.

Table 52: WM scores at Time 1

	Mean	SD	Minimum	Maximum
Listening Span	0.77	0.162	0.25	1
Digits Back L1	0.83	0.133	0.47	1
Digits Back L2	0.61	0.214	0.19	0.93
Story Recall L1	0.63	0.152	0.32	0.86
Story Recall L2	0.41	0.162	0.04	0.74

Table 53: WM scores at Time 2

	Mean	SD	Minimum	Maximum
Listening Span	0.82	0.175	0.33	1.00
Digits Back L1	0.85	0.170	0.30	1.00
Digits Back L2	0.73	0.164	0.42	0.98
Story Recall L1	0.70	0.121	0.36	0.88
Story Recall L2	0.46	0.135	0.22	0.77

Variation between measures was found, in that Digits Back L1 scores were the highest overall, and were higher than Digits Back L2; Listening Span scores were higher than Digits Back L2, and both Story Recall L1 and L2. There appears to be some interlanguage variation, in that the larger L1 scores for Digits Back and Story Recall were highly significantly different to the equivalent L2 scores (p<.01), according to Wilcoxon signed rank tests.

Differences were also found over time, illustrated in Figure 11 below, showing that all scores were higher at Time 2. Wilcoxon signed rank tests found that these differences were significant for Digits Back in L2 (p<.05), and Story Recall in L1 (p<.05), and nearly significant for Story Recall in L2 (p=.52).

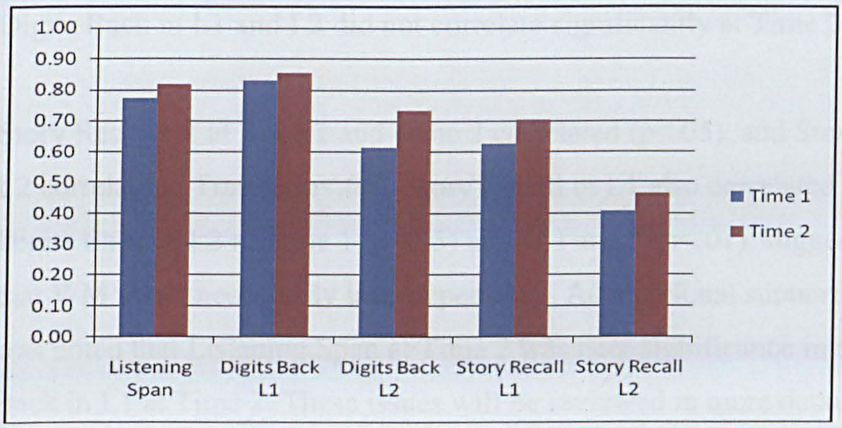


Figure 19: Comparison between WM at Time 1 and Time 2

For the L2 tests, the change over time is not necessarily surprising since it could reflect some element of increased L2 proficiency predicted to occur over the period of immersion. Although the linguistic data in this study showed little sign of improvements in the specific tasks used, the improvement shown here in L2 WM tests may be due to improvements in listening (post-test interviews with a random subset of participants found that participants' self-rated listening and comprehension skills had improved). Of more interest, in theoretical terms, is the difference in scores in the L1 and L2 versions of the same task for Digits Back and Story Recall, suggesting that adult WM may not be language-independent as argued in some of the WM literature (e.g. Osaka and Osaka 1992). Also of interest is the improvement in L1 scores for Digits Back and Story Recall, indicating that adult WM may not be as stable as standardly suggested. However, since only Story Recall in L1 was significantly different, this may be more related to the nature of the test, reflecting some element of test familiarity, rather than anything related to the nature of WM itself.

Correlational analysis was run on Digits Back and Story Recall scores to further analyse how far WM is task-dependent or not. L1 and L2 task scores at Time 1 were tested for correlations with scores at Time 2, and with each other at either time of testing. In contrast to the significant differences shown in the Wilcoxon test above, a number of significant or near-significant positive correlations were found (tables giving exact correlations are found in Appendix D). Digits Back L1 at Time 1 and at Time 2 correlated significantly ($p<.001$). Digits Back in L1 and L2 also correlated at Time 1 ($p<.001$). Digits Back L2 at Time 1 and Time 2 correlated significantly ($p<.05$), but Digits Back in L1 and L2 did not correlate significantly at Time 2.

Story Recall L1 at Time 1 and Time 2 correlated ($p<.05$), and Story Recall in L1 and L2 correlatd at Time 1 ($p<.05$). Story Recall in L1 also correlated significantly with Digits Back in L2 at Time 1 ($p<.05$) and at Time 2 ($p<.01$), suggesting the possibility that WM is not necessarily task-dependent. As additional support for this suggestion, it was noted that Listening Span at Time 2 was near significance in correlation to Digits Back in L1 at Time 2. These issues will be reviewed in more detail in the discussion chapter (chapter 6).

Given the variation between L1 and L2 versions of the Digits Back and the Story Recall tasks and variation between Time 1 and Time 2 in all the WM scores shown above, it was possible that such variation could affect the correlations with linguistic data which lay at the heart of the study's hypotheses, particularly the assumption that WM would play a role in individual variation in linguistic development. These data are now presented below.

5.4.2. Correlations between WM and oral data

Data for correlations with the oral data are presented first, showing correlations for Time 1 scores between each WM measure with the question total and question ratio scores, and copula, lexical, complex and nontarget scores, and then for Time 2 scores. However, as highlighted at the start of this section, the design of the study was intended primarily to use WM scores at Time 1 to test the assumption that WM could be implicated in individual variation in change over the period of a year's immersion (following Sagarra 2000). The correlations for WM at Time 1 with oral scores at Time 2 and with the change in oral scores are therefore also given. All correlation tables are given in full in Appendix D, but significant or near significant results and trends are summarised here.

At Time 1, Listening Span scores approached significant positive correlation with question total scores ($r=.316$, $p=.083$). Counter to expectation, Story Recall L1 was negatively correlated on all the oral scores, and was highly significant for question total ($r= -.495$, $p<.01$) and for copula questions ($r= -.542$, $p<.05$); question ratio approached significance ($r= -.338$, $p=.073$). The other WM measures showed no consistent pattern of positive or negative correlation. In other words, it seems that higher ability to repeat a story accurately in L1 correlated with lower ability to produce questions, especially simple copula questions.

At Time 2, there were no significant correlations between WM measures and oral data ($p>.1$), although there was a consistent trend of positive but trivial correlations ($r<.17$) between Story Recall in L2 and all oral measures (negative for nontarget forms, $r<-.21$). Nontarget forms were consistently negatively correlated across all WM measures, but, as stated, none of these correlations were significant. In other words, there was no significant relation between WM at Time 2 and oral scores at Time 2.

Therefore, for the oral data, the first assumption contained in Hypothesis 3 and 4, that WM was implicated in the capacity to use existing knowledge of taught question forms at fixed points in time, to access the more easily "processable" copula questions compared to lexical questions (Pienemann 1998), and to use more complex forms and fewer nontarget forms was significantly disconfirmed on at least one WM measure at the start of the period of a year's immersion, and was not supported by findings at the end of the period of immersion. Nevertheless, as discussed in section 3.1.2 above, there was some evidence of differences in linguistic development, found when the whole pool was split into three groupings according to question total scores, in that the Low-scoring group showed most improvement over time.

Therefore, to test the second assumption contained within Hypothesis 3 and 4, that WM would play a role in individual variation in linguistic improvement during immersion, WM scores at Time 1 were correlated with the oral data for Time 2. These results showed few significant correlations but some evident trends. Listening Span showed consistent positive correlations for all oral scores at Time 2 (and negative for nontarget forms), but none were significant ($p>.09$). Digits Back in L1, by contrast, showed a consistent negative trend on all measures (positive for nontarget forms), which was significant for question ratio ($r=.-454$, $p>.01$) and lexical questions ($r=-.365$, $p<.05$).[25] According to these data, Hypothesis 3 could not be confirmed or disconfirmed.

The final tests of Hypothesis 3 were to correlate WM with change on the oral measures. These findings showed that Story Recall in L1 again played a role on several oral measures, but in this analysis the role was consistently positive. Change in question total was significant ($r=.389$, $p<.05$), and change in question ratio was only just above significance ($r=.367$, $p=.05$). Change in lexical questions was also near significance ($r=.358$, $p=.057$). Listening Span showed no consistent trend, Digits Back in L1 was consistently negative on almost all measures, but by contrast, Digits Back in L2 showed a consistently positive trend (albeit non-significant, $p>.1$).

To recap the findings discussed here, those who scored higher in Story Recall in L1 at Time 1 showed significantly less ability to ask questions at Time 1, but improved the most in asking questions during immersion, especially lexical questions; they also significantly improved the most in use of complex forms and decreased the most in nontarget forms. Because this test was carried out in the L1, it is clearly language-independent. These findings finally provide some support from the oral data for Hypothesis 3, that WM may well be implicated in linguistic development during immersion. This conclusion cannot be seen as very robust, since the trend is not found across all WM measures. Nevertheless, it provides a baseline with which to compare the findings for the RT data, to which I now turn.

[25] In view of the lack of consistency across Digit Back scores with other oral measures, this is taken to be a random trend.

5.4.3. Correlations between WM and RT

5.4.3.1. WM and RT Speed

The asymmetries discussed in section 3.2.2 regarding the RT task suggested that participants were more able to access existing knowledge more efficiently, in judging taught simple question forms more quickly and accurately than either complex questions or untaught implicit ungrammatical forms, including subjacency violations, and that they judged grammatical questions more efficiently than ungrammatical questions, and object questions more efficiently than subject questions. In terms of RT times, these asymmetries were marked, and significant improvement in terms of faster speeds was found across all measures.

Turning now to test Hypothesis 3 and 4, the data were analysed to see if asymmetries and individual variation in RT times, either on arrival or after the period of a year's immersion, correlated with WM. In line with the hypotheses laid out above, lower RT times for grammatical items, short movement and long movement were expected to correlate with WM, but subjacency-constrained items were not; object questions were hypothesised to correlate more strongly than subject questions, relating to expected asymmetry in processing. Correlations were run for all RT scores for speed at Time 1 with WM measures at Time 1, then on all measures for Time 2, and between WM at Time 1 and changes in RT speed. It is noted that predicted trends would be negative – i.e. higher WM scores would correlate with lower (quicker) RT times. Full tables are given in Appendix D, and trends and significant correlations are summarised in the table below.

Table 54: Time 1 WM correlations with Time 1 RT speed

Time 1 RT measure		Story Recall L1	Story Recall L2
RT total	Correlation Coefficient	**.439(*)**	-0.106
	Sig. (2-tailed)	0.017	0.563
RT grammatical	Correlation Coefficient	**.440(*)**	-0.127
	Sig. (2-tailed)	0.017	0.49
RT ungrammatical	Correlation Coefficient	**.411(*)**	-0.111
	Sig. (2-tailed)	0.027	0.546
RT Short movement	Correlation Coefficient	**.448(*)**	-0.057
	Sig. (2-tailed)	0.015	0.757
RT Long movement	Correlation Coefficient	**.401(*)**	-0.139
	Sig. (2-tailed)	0.031	0.448
RT Subjacency	Correlation Coefficient	0.345	-0.14
	Sig. (2-tailed)	0.066	0.443
RT Object	Correlation Coefficient	**.460(*)**	-0.05
	Sig. (2-tailed)	0.012	0.786
RT Subject	Correlation Coefficient	**.427(*)**	-0.118
	Sig. (2-tailed)	0.021	0.519
N		29	32

At Time 1, there is a very clear pattern of significant correlation between SR L1 and almost all RT measures at Time 1, although note these must be interpreted in reverse: i.e. greater (slower) RT times correlated with higher WM scores. By contrast, there is a consistent pattern of negative correlation with Story Recall in L2, though non-significant and trivial. No other WM measures are significant ($p>.1$), although Listening Span, Digits Back in L1 and L2 also showed consistently positive but trivial correlations ($r<.18$).[26] This pattern of correlations with the Story Recall scores in either L1 or L2 is consistent in much of the following analysis.

The predicted asymmetries that WM would only correlate with certain forms (specifically not with subjacency-constrained items) were not robustly borne out by the

[26] This pattern of correlation between Story Recall (SR) scores in either L1 or L2 and RT measures is consistent in much of the following analyses for both RT speed and RT accuracy, so the summary tables in these two sections usually show both sets of SR scores, even when one set may not be significant, to allow for direct comparisons of how the trends of correlation shift between the two sets.

findings, in that the pattern found for Story Recall in L1 was consistent across all subdivisions, and the correlation for subjacency was close to significant (p<.07). In other words, higher ability to recall a story accurately in L1 was significantly associated with taking longer to judge the RT items across all types of questions. The similar ability to recall a story accurately in L2 was consistently associated with faster RT judgements, but the correlations are too trivial to draw any statistically valid conclusions. For RT speed at Time 2, Story Recall again was the only test to show any consistent pattern of correlation, as shown in the table below. But here the pattern is reversed, in that it is Story Recall in L2 which showed strong positive correlations with RT speeds at Time 2 on all measures.

Table 55: Time 2 WM correlations with Time 2 RT speed

Time 2 RT		Story Recall L1	Story Recall L2
RT	Correlation Coefficient	.194	**.528(**)**
	Sig. (2-tailed)	.288	.002
RT grammatical	Correlation Coefficient	.254	**.484(**)**
	Sig. (2-tailed)	.160	.005
RT ungrammatical	Correlation Coefficient	.184	**.498(**)**
	Sig. (2-tailed)	.312	.004
RT Short movement	Correlation Coefficient	.171	**.442(*)**
	Sig. (2-tailed)	.350	.011
RT Long movement	Correlation Coefficient	.228	**.495(**)**
	Sig. (2-tailed)	.209	.004
RT Subjacency	Correlation Coefficient	.192	**.590(**)**
	Sig. (2-tailed)	.291	.000
RT Object	Correlation Coefficient	.158	**.459(**)**
	Sig. (2-tailed)	.389	.008
RT Subject	Correlation Coefficient	.190	**.446(*)**
	Sig. (2-tailed)	.298	.011
N		32	32

The significant correlation coefficient values found for Story Recall L1 at Time 1 compared to Story Recall L2 at time 2 with RT speeds is illustrated for clarity in the figure below.

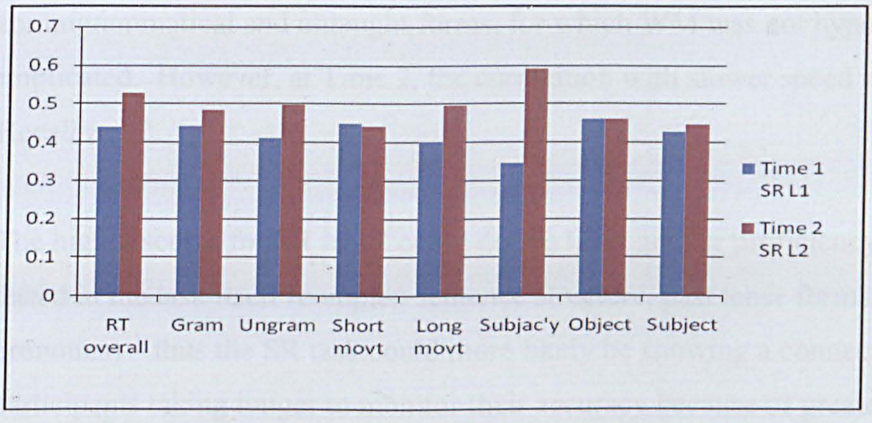


Figure 20: Correlations between Story Recall L1 at Time 1 with RTs at Time 1 compared to correlations between Story Recall L2 at Time 2 with RTs at Time 2

Comparing the patterns of correlations for these data at Time 1 and Time 2, there is mixed evidence to support the hypothesised asymmetric relation between WM and specific submeasures. Expected patterns of correlation were that faster speeds on short movement and long movement would correlate with WM (i.e. would show negative correlations), but subjacency-constrained items would not (i.e. no significant positive or negative correlation). Grammatical items and object questions were hypothesised to correlate more strongly than ungrammatical and subject questions, relating to expected asymmetry in processing (i.e. grammatical items and objects would show higher negative correlations than ungrammatical items and subjects).

The data however do not bear these assumptions out. As illustrated clearly, the pattern of correlations was consistently positive and all around between .4 and .6 strength of Spearman's rho. At Time 1, grammatical items, short movement and objects were, against prediction, slightly higher than the ungrammatical items and subjects. At Time 2, the expected asymmetry was found, but only ungrammatical items and Long movement showed a marked difference. Short movement and objects showed almost no difference between strength of correlations at Time 1 compared to Time 2. Contrary to predictions, subjacency showed significant correlations at both times, most strongly at Time 2.

These data emphasise the conclusion drawn at Time 1, that higher ability to recall a story accurately is associated with slower RT judgements across all measures, not only for the explicit taught forms which WM was, in fact, hypothesised to facilitate, but also for ungrammatical and untaught forms, for which WM was not hypothesised to be implicated. However, at Time 2, the correlation with slower speed was found in Story Recall in L2.

The higher scores for SR in L2 could derive from greater proficiency on the grammar tested in the task itself (complex sentence structure, past tense forms, gender-specific pronouns) – thus the SR task could more likely be showing a connection between participants taking longer to monitor their accuracy because of greater L2 proficiency, rather than greater WM capacity. However, the similarity of patterns between Story Recall in L1 and Story Recall in L2 with slower speeds suggests a language-independent explanation is more sustainable, and matches significant correlations found between the two Story Recall tasks at Time 1 ($r=.424$, $p<.05$).

The unexpected patterns of correlation between WM and slower RT speeds seemed to counterindicate the fundamental assumption of this study that WM would facilitate improvement over the period of immersion, seen in correlations between WM at Time 1 and improvements in RT speeds at Time 2. In order to check this, correlations were run between WM at Time 1 and RT speeds at Time 2, and between WM at Time 1 and change in RT by Time 2. As before, neither Listening Span nor Digits Back tasks showed any significant correlations with RTs ($p>.5$). Consistent significant correlations with RT at Time 2 were found for Story Recall in both L1 and L2, and between and with change in RT and Story Recall in L2, as summarised in the tables below (full tables showing all correlations are in Appendix D).

Table 56: Correlations between Story Recall at Time 1 with RT speed at Time 2

RT at Time 2		Time 1 Story Recall L1	Time 1 Story Recall L2
RT Time 2	Correlation Coefficient	**.581(**)**	**.389(*)**
	Sig. (2-tailed)	.001	.028
RT gram Time 2	Correlation Coefficient	**.624(**)**	**.389(*)**
	Sig. (2-tailed)	.000	.028
RT ungram Time 2	Correlation Coefficient	**.514(**)**	**.375(*)**
	Sig. (2-tailed)	.004	.034
RT Short movement	Correlation Coefficient	**.608(**)**	.327
Time 2	Sig. (2-tailed)	.000	.068
RT Long movement	Correlation Coefficient	**.580(**)**	**.420(*)**
Time 2	Sig. (2-tailed)	.001	.017
RT Subjacency	Correlation Coefficient	**.490(**)**	.324
Time 2	Sig. (2-tailed)	.007	.071
RT Object questions	Correlation Coefficient	**.574(**)**	.304
Time 2	Sig. (2-tailed)	.001	.090
RT Subject questions	Correlation Coefficient	**.631(**)**	**.362(*)**
Time 2	Sig. (2-tailed)	.000	.041
N		29	32

Table 57: Correlations between Story Recall at Time 1 with Change in RT speed

		Story Recall L1	Story Recall L2
Change RT	Correlation Coefficient	-.001	**.402(*)**
	Sig. (2-tailed)	.997	.022
Change gram RT	Correlation Coefficient	.032	.346
	Sig. (2-tailed)	.870	.052
Change ungram RT	Correlation Coefficient	-.002	**.388(*)**
	Sig. (2-tailed)	.991	.028
Change Short movt RT	Correlation Coefficient	.076	.300
	Sig. (2-tailed)	.696	.096
Change Long movt RT	Correlation Coefficient	-.003	**.382(*)**
	Sig. (2-tailed)	.987	.031
Change Subjacency RT	Correlation Coefficient	.029	**.409(*)**
	Sig. (2-tailed)	.883	.020
Change Object RT	Correlation Coefficient	-.066	.220
	Sig. (2-tailed)	.736	.226
Change Subject RT	Correlation Coefficient	.063	**.425(*)**
	Sig. (2-tailed)	.745	.015
N		29	32

The correlations shown in Table 56 above for scores for WM at T1 and RT Speed at T2 showed that SR L1 was highly significant on all measures (p<.01), but also that SR L2 was near or at significance for all measures (p<.05).

For change in RT speed (Table 57), a very clear pattern of correlation with Story Recall in L2 is seen, with significant or near significant correlations (p<.05), apart from change in object speed (p>.2). Story Recall in L1, however, showed no clear pattern negatively or positively with change in RT speed, which is surprising, given the strong correlations found between SR L1 and slower speeds at both T1 and T2.

To recap the findings discussed here, the pattern of positive correlations for Story Recall in both L1 and L2 with RT speed remains consistent across all correlational analyses shown above for on Time 1 measures, Time 2 measures and on Change in RT speed measures. This suggests that greater WM capacity may lead to longer, more conscious reflection about RT judgements, despite the instruction given to participants to respond as instinctively and quickly as possible. Greater capacity to use WM as a "monitoring" or "hypothesis-testing" workspace for judgements of grammaticality could therefore make RT speeds slower, both before and after immersion (even though RT speed in itself improved after immersion).

In addition, WM capacity seems to correlate similarly across all subdivisions. The predictions that WM would correlate positively with some types but not others (i.e. not with subjacency) do not appear to be borne out.

In sum, Hypotheses 3 and 4 about the implication of WM in linguistic development as measured in RT speed are disconfirmed – rather than facilitate faster RT speeds on certain submeasures, WM seems to be implicated in slower, more monitored RT speeds, and this pattern is reflected across all submeasures.

I turn now to see if these patterns are reflected in the findings for RT accuracy.

5.4.3.2. WM and RT Accuracy

Correlations between WM and accuracy scores are presented here, in line with Hypotheses 3 and 4, that WM correlates with greater accuracy in RT judgements, particularly in terms of asymmetries between submeasures, and with individual variation in improvements in accuracy over the period of immersion. In view of the lack of significant correlations between speed and accuracy described in section 5.3.3, it was potentially unlikely that the significant correlations found for Story Recall in L1 and L2 with slower speeds at Time 1 and Time 2 would be found. In addition, given the lack of significant change in accuracy in the RT task, it could be assumed that there would be no robust basis to draw any consistent correlations between WM and RT accuracy, to support Hypothesis 3 and 4.

Running the analysis first for WM scores at Time 1 with RT accuracy scores at Time 1, summarised in the table below (full tables are presented in Appendix D), there is, as expected, a mixed pattern of evidence. Story Recall in L1 is the only measure which is significantly correlated, showing positive correlations with RT scores for ungrammatical accuracy and subjacency ($p<.05$), but a negative correlation for LM accuracy ($p<.05$). Listening Span and Digits Back in either L1 or L2 do not show any significance ($p>.14$).

Table 58: Correlations between WM at Time 1 and RT Accuracy at Time 1

Accuracy		Story Recall L1	Story Recall L2
Accuracy	Correlation Coefficient	.192	-.032
	Sig. (2-tailed)	.319	.861
Gram accuracy	Correlation Coefficient	-.182	.021
	Sig. (2-tailed)	.345	.911
Ungram accuracy	Correlation Coefficient	**.368(*)**	-.114
	Sig. (2-tailed)	.049	.536
SM accuracy	Correlation Coefficient	.247	-.050
	Sig. (2-tailed)	.196	.785
LM accuracy	Correlation Coefficient	**-.439(*)**	-.155
	Sig. (2-tailed)	.017	.396
Subjacency	Correlation Coefficient	**.402(*)**	.014
	Sig. (2-tailed)	.030	.940
Object	Correlation Coefficient	.093	-.196
	Sig. (2-tailed)	.633	.283
Subject	Correlation Coefficient	-.160	-.085
	Sig. (2-tailed)	.406	.645
N		29	32

Turning now to correlations between WM at Time 2 and RT Accuracy measures at Time 2, there were no significant correlations for any RT measures with any WM measures ($p>.05$), unlike those found between Story Recall in L2 with slower RT speeds. The closest to significance was found between Digits Back in L2 with Short movement accuracy ($r=.345$, $p=.053$).

There seems, therefore, to be no evidence to support the assumption that greater WM capacity is implicated in higher RT Accuracy scores at either Time 1 or Time 2. However, as with the other analyses, correlations were also carried out to check for patterns of correlation between WM at T1 and Accuracy at Time 2, and between WM at Time 1 with Change in Accuracy scores. For WM at T1 and Accuracy measures at

T2, there were a number of significant or near-significant correlations summarised in the table below. Digits Back in L2 showed a positive correlation with short movement accuracy (p<.05) and approaching significant positive correlation with ungrammatical accuracy (p=.08). However, there was wider evidence of significant negative correlations; Digits Back in L1 correlated negatively with subject accuracy (p<.05); Story Recall in L1 correlated negatively for grammatical accuracy (p<.05), and long movement accuracy (p<.05). Story Recall in L2 and Listening Span showed no significant or consistent patterns (p>.1). Subjacency correlated positively with both Digits Back in L1 and L2, and with Story Recall in L1 and L2, but all correlations were non-significant (p>.1).

Table 59: Correlations between WM at Time 1 and Accuracy at Time 2

Accuracy Time 2		Digits Back L1	Digits Back L2	Story Recall L1
Grammatical	Correlation Coefficient	-.287	-.029	**-.470(*)**
	Sig. (2-tailed)	.117	.878	.010
Ungrammatical	Correlation Coefficient	.097	**.321**	.352
	Sig. (2-tailed)	.605	.083	.061
Short movement	Correlation Coefficient	-.073	**.363(*)**	.116
	Sig. (2-tailed)	.696	.049	.548
Long movement	Correlation Coefficient	-.255	-.147	**-.437(*)**
	Sig. (2-tailed)	.167	.438	.018
Subjects	Correlation Coefficient	**-.363(*)**	.050	-.236
	Sig. (2-tailed)	.045	.795	.217
N		31	30	29

For correlations between WM at Time 1 and Change in accuracy scores, the WM ratio scores showed no significant correlations on any measures (p>.05). There were near-significant findings for Digits Back in L2 with both Change in Accuracy (r= .348, p=.06) and change in Short movement (r= .334, p=.07). Digits Back in L1 and Story Recall in L1 yielded non-significant trivial but consistently negative correlations. Listening Span revealed no consistent patterns.

Given the lack of consistent significant improvements in RT accuracy scores over time, as seen in Table 45 in section 5.3.2.2 earlier, the inconsistent patterns of correlation with WM and lack of significance is not unexpected. As shown in the discussion on RT accuracy, overall accuracy scores decreased, arising particularly from significant decreases for ungrammatical and subjacency-constrained items ($p<.05$). Only grammatical items increased in accuracy significantly ($p<.01$) and subject questions improved but nonsignificantly. The correlations with WM scores do not, however, reflect this pattern of improvement, in that the expected correlation between higher WM scores and change in accuracy on grammatical items or subject items is not found – in fact there is evidence of significant negative correlation between Story Recall in L1 and change in grammatical accuracy scores.

To conclude this discussion of WM correlations with RT accuracy scores, the mixed patterns of positive and negative correlations across WM measures and across all RT accuracy scores do not provide clear evidence in support of Hypothesis 3 or 4.

As with the RT speed correlations shown in section 5.4.3.1 above, there is a very consistent pattern of correlation between Story Recall scores and the RT accuracy measures, particularly Story Recall in L1. It may be that higher WM capacity, such as the ability to recall a story accurately, simply makes it more likely that the response to the RT items is both slower and less accurate, even for grammatical items, which are expected to be known explicitly from the taught input, and were predicted to be more easily retrieved, and implicated in greater WM capacity. It may be that greater WM capacity drives learners to question their own judgements more hesitantly, especially over the period of immersion tested here, in that learners' underlying L2 knowledge may be being restructured following a kind of "U-shaped" trajectory, as is commonly found in child acquisition. Judgements could thus become less accurate during immersion than at the start, before learners' underlying knowledge eventually begins to be restructured to provide more accurate, faster intuitions, but this may require longer immersion than was tested here. It may simply be that, for a number of reasons, the WM correlations shown here are little more than random patterns.

To check further whether the general patterns found above were random or obscured any underlying patterns of individual variation, particularly in view of the between-group differences found in RT accuracy (section 5.3.2.2 above), further analysis was run to check the highest scoring individuals at Time 2 who could be deemed to have acquired wh-movement over the period of the study. In terms of acquisition (e.g. Vainikka and Young-Scholten 1994), 60% accuracy can be taken to show acquisition, and fourteen out of thirty-two participants scored 41 or above on their raw accuracy score. All participants' scores on change in accuracy were grouped for high, mid and low change, and these fourteen participants were checked against this group rating. Of these fourteen participants, three were low change, five were mid change and six were high change as summarised in the table below:

Table 60: Most accurate participants by Time 2

ID	Accuracy score at Time 2	%	Change in Accuracy by group
SHE	**56**	82.35%	**High**
VIO	**52**	76.47%	**High**
TRD	**50**	73.53%	**High**
MAG	**50**	73.53%	**High**
CEC	**48**	70.59%	**High**
SZU	**41**	60.29%	**High**
LUJ	48	70.59%	Mid
ANN	46	67.65%	Mid
BOZ	46	67.65%	Mid
ERI	45	66.18%	Mid
STE	45	66.18%	Mid
DAV	52	76.47%	Low
CIN	46	67.65%	Low
WEI	41	60.29%	Low

This means that three who scored highly at Time 2 were already high scorers from Time 1, and showed low change (DAV, CIN and WEI). However six who scored highly at Time 2 were in the high change group, suggesting these participants might fulfil the expected assumptions of linguistic improvement or acquisition underpinning the whole study.

Between-group analysis of WM scores (Kruskal-Wallis tests) on all participants split by rating of RT accuracy at Time 2 showed no significant between-group differences on either WM or change in accuracy scores – i.e. the highest-scoring group at Time 2 were not significantly different to the rest in the distribution of their WM scores, nor in the pattern of change in accuracy scores ($p>.05$). The full results of the Kruskal-Wallis tests are shown in Appendix D.

Examining the correlation results just on the fourteen highest scorers showed that, at Time 2, there were a number of significant correlations between WM measures and different accuracy measures, summarised in the table below. For the first time in any of the correlations run hitherto, Listening Span shows a significant positive correlation with overall accuracy ($p<.05$), and with short movement ($p<.05$). Story Recall in L1 shows a significant positive correlation with ungrammatical accuracy ($p<.01$). However, Story Recall in both L1 and L2 show significant negative correlations with grammatical accuracy ($p<.05$), and significant or near negative correlations with long movement. Full correlation tables showing high, mid and low groups are shown in Appendix D.

Table 61: Correlations between WM and highest scores for RT accuracy, Time 2(using groups split by Accuracy at Time 2)

High group		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Accuracy	Correlation Coefficient	**.618(*)**	.234	.519	.279	-.097
Time 2	Sig. (2-tailed)	.025	.441	.084	.356	.742
Grammatical accuracy	Correlation Coefficient	.109	-.277	.124	**-.596(*)**	**-.550(*)**
Time 2	Sig. (2-tailed)	.722	.359	.702	.032	.041
Ungrammatical accuracy	Correlation Coefficient	.345	.364	.274	**.692(**)**	.347
Time 2	Sig. (2-tailed)	.248	.222	.389	.009	.224
Short movement	Correlation Coefficient	**.624(*)**	-.129	.522	.357	-.156
Time 2	Sig. (2-tailed)	.023	.674	.082	.231	.594
Long movement	Correlation Coefficient	-.227	-.100	-.191	**-.670(*)**	-.516
Time 2	Sig. (2-tailed)	.455	.746	.552	.012	.059
N		14	14	14	14	14

There were no significant correlations for this highest-scoring group between WM and changes in accuracy ($p>.07$).

For the Low group ($n=7$), who all scored at chance or below, there were no significant correlations ($p>.05$) with accuracy at Time 2 or for change in accuracy on any measures, other than for subjacency at Time 2 ($r=-.912$, $p<.05$).

The Mid group ($n=11$), who scored between chance and 41 (out of a total of 68) at Time 2, showed some significant negative correlations between WM and accuracy at Time 2, as shown in the table below. Story Recall in L1 significantly negatively correlated with grammatical accuracy ($p<.05$), Story Recall in L2 significantly correlated with short movement ($p<.05$), and Digits Back in L2 significantly correlated with subject accuracy ($p<.05$). There were no significant correlations between WM and change in accuracy scores.

Table 62: Correlations between WM and mid-group for RT accuracy at Time 2

		Digits Back L2	Story Recall L1	Story Recall L2
Grammatical accuracy Time 2	Correlation Coefficient	-.315	**-.726(*)**	-.096
	Sig. (2-tailed)	.345	.018	.779
Short movement Time 2	Correlation Coefficient	.148	.143	**-.608(*)**
	Sig. (2-tailed)	.664	.694	.047
Subject questions Time 2	Correlation Coefficient	**-.644(*)**	-.234	-.373
	Sig. (2-tailed)	.033	.514	.259
N		11	10	11

These findings provide some evidence that WM was potentially implicated more positively for the highest-scoring group than for the lowest scoring group.

To confirm that Story Recall in L1 was the most consistent test relating to the RT Accuracy scores, the whole pool was divided into three groups, split by scores on each WM measure (Low N=10, Mid N=10, High N=9). Kruskal-Wallis tests of between-group difference on Accuracy at Time 2 showed that Story Recall in L1 was the only test to show any significance (chi-square 8.792, p<.05). Kruskal-Wallis tests on Change in accuracy showed no significance (p>.05). The mean scores for Accuracy at Time 2 for each of the three groups split by Story Recall in L1 score revealed that the between-group significance derived from the mid group showing the least accuracy compared to the other two groups. Mean accuracy for the mid group was 33.20 (out of 68), compared to the low group (mean accuracy 43.20) and the high group (mean accuracy 43.22).

In order to test the conclusion that Story Recall in L1 was the most statistically significant test for Accuracy at Time 2, the whole pool of participants was split by Story Recall (into low, mid, high groups, using equal percentile cutpoints). Kruskal-Wallis tests for between-group differences for the different accuracy submeasures showed that only ungrammatical and subjacency scores were significantly different between the groups (p<.05). This corroborated the general finding discussed earlier that Story Recall in L1 was implicated in accuracy in subjacency across the whole pool but not in the other accuracy submeasures. Comparison of mean subjacency scores between the different groups split by Story Recall confirmed that the high group had the highest subjacency mean accuracy scores. The high group for Story Recall in L1 scored 7.78 mean accuracy (out of 12) compared to 5.10 for the low group and 3.8 for the mid-group. At Time 2, the high Story Recall group were again more accurate on subjacency than the other two groups, though the between-group differences were not significant (full details are given in Appendix D).

Although there was no obvious evidence from correlational analysis of a trade-off just between RT speed and accuracy (p>.1), shown in section 5.3.3 above, between-group analysis suggested that those with higher Story Recall capacity did take longer but scored more accurately, particularly on subjacency. Kruskal-Wallis tests found significant between-group differences on subjacency speeds at Time 1, and comparison of mean RT speeds confirmed that the high Story Recall group showed slower (greater) mean speeds than the other two groups. The high group was also slower at Time 2, although the between-group differences were not significant (see Appendix D for full details).

Thus it seems that there may have been some kind of interaction between greater Story Recall capacity, slower RTs and greater accuracy, yielding the inference that those with greater Story Recall capacity were not only taking longer but were achieving greater accuracy by doing so.

In sum, there are some significant correlations between some WM measures which support the hypotheses of this study, including positive significant correlations between the highest scorers for Story Recall in L1 with their scores for subjacency accuracy and response times, and between Listening Span and Accuracy for a subgroup of highest scoring participants at Time 2. However, there are also other consistently negative correlations between RT scores and Story Recall, across all participants, which counter-indicate or disconfirm the hypotheses.

5.4.4. Posthoc correlations between WM and oral fluency

In order to complete the analysis of all the data discussed in this chapter, including the posthoc analysis of the oral fluency data, WM measures were also correlated with the scores for increase in type-token ratio and decrease in hesitancy phenomena (see discussion in section 5.3.1.3). It was hypothesised, in view of some evidence of positive correlation between WM capacity and oral fluency found in Temple (1997), Fortkamp (1999) and O'Brien et al (2006), that the improvement in type-token ratio and decline in hesitancy phenomena would correlate with WM. Non-parametric correlations were run on these measures (see Appendix D). However, there were no consistent patterns of significant correlations between greater WM capacity and lower levels of repair, or with higher levels of improvement (correlations were above $p>.1$, apart from Digits Back in L2 and type-token ratio at Time 2). This finding is similar to Mizera's (2006) study in which he failed to find significant correlations between WM and oral production in L2 Spanish learners.

5.4.5. Conclusion

In conclusion, it was not possible to find robust significant evidence to confirm all four research hypotheses driving this study. In terms of linguistic development, efficient use of accurate wh-questions appeared to remain problematic and did not appear to significantly change during immersion as measured by oral and RT tasks, other than in faster RT speeds. There were some descriptive differences in line with the predicted asymmetries that instructed learners would be more efficient in perception/production of simple questions, grammatical questions and object questions, compared to complex questions, especially ungrammatical items, and subject questions, but these differences were not always significant. Hypotheses 1 and 2 are therefore not confirmed.

There was no consistent pattern of significant correlations between any changes and WM, apart from the correlations found between SR in L1 and L2 and slower times in the RT task; these significant correlations were spread across all measures, disconfirming Hypothesis 3 and 4, that WM would be positively correlated with faster speeds on most measures apart from implicit untaught ungrammatical forms including subjacency-constrained items.

In addition, post-hoc analysis of the oral data for improvements in fluency did not show any pattern of consistent correlations with WM.

I will turn in the next chapter to discuss why this might be so, revisiting the original theories and models for acquisition which drove the design of this study, such as the distinctions between explicit and implicit knowledge, or between learned knowledge and acquired competence, how input is argued to increase strength of underlying representations, as argued in the MOGUL model, and how far it is possible to argue for the role of WM in the process of SLA. I also reassess how far other methodological factors could affect the data, such as test reliability for the oral and RT tasks, and the WM tasks, and if other factors need to be reconsidered, even where these were assumed to be reasonably homogenous, such as participants' biodata.

Chapter 6: Second study - Evaluation and discussion

6.1. Introduction

The second study tested four hypotheses predicting that there would be clear patterns of variation found in linguistic development of L2 English wh-movement by advanced adult Chinese speakers during immersion in the UK, and that WM would play a important role in explaining individual variation in that development. The four hypotheses were:

Hypothesis 1: Instructed Chinese learners of English will show asymmetries in acquisition and use of question forms, tested through oral output and timed grammaticality judgments, measured as:

a) greater use of targetlike simple questions compared to complex questions in oral output;
b) faster speed and accuracy in timed judgments on simple questions than complex questions, especially untaught implicit forms, including subjacency violations;
c) faster speed and accuracy in timed judgments on grammatical questions compared to ungrammatical questions and object questions compared to subject questions.

Hypothesis 2: These learners will improve over time in their knowledge and use of question forms (subject to the asymmetries noted in 1) when they are exposed to enriched input through increased primary linguistic data from native speakers in an immersion setting.

Hypothesis 3: WM capacity is implicated in participants' ability to access existing knowledge of taught question forms more efficiently, in that greater WM capacity would correlate with individual differences in perception/production of targetlike questions (other than for subjacency-constrained items) on both linguistic tasks, and with variation in rates of improvement over time during immersion.

Hypothesis 4: WM capacity is not implicated in participants' capacity to acquire untaught implicit subjacency constraints measured on a timed grammaticality judgement task.

However, as shown in the previous chapter, it was not possible to find robust evidence to confirm all four research hypotheses driving this study.

6.2. Key conclusions

In the oral task, participants appeared to be more efficient in accessing knowledge of the simplest possible question forms, preferring copula to lexical questions, and producing few complex embedded questions at both Time 1 and Time 2.

Post-hoc analysis of nontarget oral forms suggest that problems in targetlikeness derive from difficulties in verb raising, with most participants showing evidence of being at stage 3 in Pienemann's hierarchy of stages of questions, showing difficulties in acquiring the full range of features at TP (Hawkins 2001; Vainikka and Young-Scholten 2005). Tense marking was highly optional and tended to be marked on verbs in-situ or on a default *be* marker. An implicational hierarchy of nontargetlike wh-questions prior to targetlike production is suggested, based on processing of lexical knowledge. The progression starts with use of chunks and omission of verbs or use of bare verbs without tense marking, then optionality between default use of *be* and in-situ tense marking; once default *be* can be separated from *do* as a marker of verb raising, targetlike wh-movement can be found.

In the RT task, there was evidence of predicted asymmetries that instructed learners would be more accurate and faster in judging simple short-movement questions, grammatical questions and object questions, compared to complex long-movement questions, ungrammatical questions, subject questions, and especially subjacency-constrained questions, but these differences were not always significant.

Notably, immersion did not have the expected effect of triggering marked linguistic development, at least according to group scores for targetlike questions in the oral production task, or RT accuracy, although RT speeds got significantly faster. The assumption was drawn that improvement in speed was due to improved automaticity.

To check this assumption, the RT data were compared in a post-hoc analysis of fluency measures in oral data. The post-hoc analysis of oral fluency, as measured by type-token ratio and repairs and filled pauses, improved significantly. If oral fluency and faster RT speeds can both be taken as evidence of greater automaticity (Towell and Dewaele 2005), then it seems clear that the evidence of significantly greater automaticity will have been the result of immersion.

Assessing the role of WM in linguistic development, five measures were used to test for correlations with linguistic data: Listening Span in L2 (following methodology similar to Harrington and Sawyer 1992); Digits Back in L1 and L2 (following standard test procedures used by Caplan and Waters 1996); Story Recall in L1 and L2 (used by Fehringer and Fry 2007). Of these, Story Recall (in L1 and L2) showed the most consistent patterns of correlation with linguistic scores on both oral and RT tasks (as had been found by Fehringer and Fry 2007). Story Recall in L1 correlated below or near significance with improvement in question total ($p<.05$) and with improvement in lexical questions ($p=.057$). There was a clear pattern of significant correlations between Story Recall (in L1 and L2) and slower RT speeds ($p<.05$) at both times of testing. These significant correlations were stronger on explicit forms at Time 1, but were spread consistently across all measures at Time 2. By contrast, there were no positive correlations found on any WM measure with RT accuracy, at either Time 1 or Time 2, or with change in accuracy, apart from a significant positive correlation between Story Recall in L1 and accuracy on subjacency-constrained forms at Time 1. The prediction of an asymmetric relation between WM and different linguistic measures, in which explicit taught forms were assumed to correlate with WM but not implicit nontaught subjacency-constrained items, was therefore not borne out.

The findings strongly indicate that acquisition of wh-movement (and, by implication, verb raising) as defined in terms of efficient, nonmonitored (or implicit) targetlike perception/production (Herschensohn 1999; Paradis 2009) was not in place by the end of Time 2. Optionality in the oral data and doubt over the RT judgements are the key features of the data reported here. Clearly, asking questions is found to be a difficult task by the Chinese participants in these studies, even after 11 months' immersion. As one participant said, when her test session finished at Time 2, "I still can't questions".

Immersion, for a year, at least, therefore does not seem to provide the kind of rich exposure to primary linguistic data which was expected to be sufficient to trigger changes in underlying knowledge (contra Howard 2006). Instead the improvements in RT speed and oral fluency are in line with other evidence that the kind of immersion experienced through study abroad tends to lead to improvements in greater automaticity (Segalowitz 2003) and better oral fluency (Freed et al 2003, O'Brien et al 2006), rather than changes in morphosyntactic accuracy. In other words, immersion indeed affected participants, but did so in terms of processing/easier retrieval of existing knowledge, rather than triggering acquisition of new linguistic knowledge. This was borne out in terms of participants' subjective assessment of progress - judging from informal post-hoc interviews, they were aware they were better able to comprehend oral input and could read and write more quickly, but they did not feel their grammatical accuracy had improved nor their spoken accuracy or overall confidence in speaking fluently.

The lack of material development in linguistic knowledge during immersion undermined the central investigation of this study of how WM might correlate with individual variation in development during immersion. The significant correlations which were found between WM and the linguistic data seemed to yield contradictory results. The positive correlation between Story Recall and improvement in the oral task suggests that WM could be implicated in the development of greater morphosyntactic accuracy in online production. By contrast, the positive correlations between Story Recall and slower RT speeds, but not with greater RT accuracy, suggest that greater WM capacity may primarily facilitate promote slower, more reflective or more hesitant responses. It also seems, from the data, that WM is not specifically implicated in processing explicit learned knowledge compared to implicit acquired knowledge.

The central question of how far WM capacity is implicated in acquisition and use of L2 thus remains unresolved here.

6.3. Potential reasons for lack of linguistic development

There are a number of possible reasons why the expected outcomes for linguistic development in oral and RT measures were not found.

In terms of morphosyntactic development, it had been assumed that development would follow predictable patterns in line either with theories of acquisition of L2 features or with processing-based accounts. Different generative theories of feature-based acquisition, committed to Universal Grammar constraints playing a role in adult L2A, would assume that L2 features should be acquired – in this case, strong [+wh], strong [Q] and strong [+Tense], driving wh-movement and verb raising. However, the different theories predicted different effects of L1 transfer. Structure-building accounts (such as Organic Grammar) assumed minimal L1 transfer; other accounts (such as Full Transfer/Full Access) assumed maximal transfer. The Representational Deficit Hypothesis (Hawkins and Liszka 2003) goes further to assume that features not instantiated in the L1 cannot be acquired by adult L2 learners. Other accounts predict difficulties at the morphology/syntax interface, where syntactic features may well be acquired, but production of surface inflectional morphology would be impaired, in line with the Missing Surface Inflection Hypothesis (Prévost and White 2000; Lardiere 2008, 2009). The use of both oral production and grammaticality judgement data in this study was intended to shed light on how far non-targetlikeness could be argued to be based on syntactic deficit or difficulties at the morphological interface.

The data showed a disparity between targetlikeness in the oral task compared to the RT task. Nearly half the group reached 60% accuracy or more on the grammaticality judgement RT task, on the one hand, but there was wide optionality in verb raising and lack of lexical questions shown by the whole group in the oral data, on the other, as well as wide evidence of slow RT speeds. The evidence of some level of successful acquisition would argue against the Representational Deficit Hypothesis; the evidence of L1 transfer in lack of tense marking and verb raising could be taken to argue against minimal transfer accounts such as Organic Grammar, although it could simply be evidence of participants' underlying competence remaining at the VP stage, representing early stages of Organic Grammar regardless of L1. The disparity between the oral and RT data suggests that most of the difficulty could arise in production of the surface morphology, in line with the Missing Surface Inflection Hypothesis (MSIH). However, the evidence of wide optionality and verb marking in situ overrides the predictions of the MSIH, suggesting that other forces are also at play. The evidence also of hesitancy in the oral task and slow RT speeds drives the conclusion that an account based only on acquisition of feature strength is insufficient for a full understanding of the data.

In terms of processing-based accounts, models such as Pollard and Sag's Head-driven Phrase Structure Grammar and Culicover and Jackendoff's Simpler Syntax Hypothesis, suggested that shorter questions would be processed more easily than long distance questions, which was indeed found; but again, the models cannot fully account for the optionality and asymmetries found in the data, especially the discrepancy between the oral and RT levels of performance. Taking Pienemann's Processability Theory, it was assumed that output would show a clear improvement up the hierarchy of question stages depending on what type of underlying stage of phrase or clause structure had been established in the underlying grammar. This was not clearly shown in the data, and again the optionality and heavy reliance on default *be* in the oral data suggested that there was something else needed to explain the data other than stage of processability. Processability Theory would also not be able to give precise explanations for some of the asymmetries found in the RT data, especially between subject and object questions, and grammatical and ungrammatical questions, nor for the disparity between faster speeds but no improvement in accuracy.

Truscott and Sharwood Smith's MOGUL model appears to be most in keeping with the data here. The concept of constrained and predictable feature-based development, in which L1 and L2 features operate at competing levels of strength of activation, could account for the evidence that participants were able to show offline awareness of L2 features (as shown in the RT accuracy ratings by nearly half the group) while also showing online difficulty in combining all the features needed for successful wh-movement, tense marking and verb raising in oral production. In Truscott and Sharwood Smith's model of "Acquisition by Processing", individual differences in processing capacity, and differences in the processing of different types of question forms, may well explain the optionality and hesitancy found in the data.

An additional issue in explaining the apparent discrepancy in performance between the oral and the RT task may lie in different processing strategies between online oral performance and grammaticality judgements using written prompts (Murphy 1997). This discrepancy may well reflect the use of different knowledge sources as hypothesised in this research, but can confound the discussion in SLA over whether data are a measure of acquisition or use (or both). As outlined in the literature review (Chapter 2), there have been concerns about the use of grammaticality judgements compared to oral data, in that online oral data are intended to tap implicit competence while grammaticality judgements may more consistently tap other sources such as explicit metalinguistic awareness (Birdsong 1992; Bialystok 1994; Sorace 2003). In addition, there is the issue of how far nontargetlike data reflect issues in processing and performance, or in acquisition. This potential confound was controlled for in the Reaction Time task, to some extent, because the use of grammatical taught forms was carefully balanced against untaught ungrammatical and implicit-constrained forms. The patterns of responses did indeed show some degree of greater accuracy on the more explicitly taught forms that could have been derived from metalinguistic knowledge. However, there was counter-evidence that response times for these forms were faster, suggesting some element of greater automaticity on these forms, rather than slower metalinguistic reflection. In addition, evidence of accuracy on ungrammatical and untaught forms suggests that underlying implicit linguistic knowledge was being tapped, even if explicit knowledge also played a role to some degree.

There were also specific methodological issues with the oral task itself, arising from the heavy reliance on copula forms used throughout the task by all participants. It is suggested that production relies on pre-fabricated chunks (Myles 2004) which, in the case of frequently drilled question forms, as seen here, may or may not have been stored in non-target form, depending on the original learning situation. Chunks are understood to be retrieved non-consciously, without much conscious monitoring for non-target forms, since this is a costly strategy (Pienemann 1998; Jackendoff 2002; Paradis 2009). I infer that participants are aiming for a pragmatic approach in responding to the task – that if they can manage to get their meaning across, target-like forms can be too difficult to control, particularly given the instruction they had been given to ask "as many questions as you can in five minutes". This inference is supported by comments from participants who were aware of the time limit and sometimes felt they had often asked all the questions they needed to ask well before the five minutes were completed, leading to a degree of prompting of questions for questions' sake. This pragmatic strategy may be reflected in the wide range of actual output produced in the time allowed (e.g. from 23 to 64 utterances in total, and, out of these, from four to 26 accurate questions, at Time 2). The oral task may not therefore have been a reliable measure to tap linguistic knowledge, and further research using a wider range of oral elicitation tasks would help to clarify this possible difficulty.

6.4. Issues arising from lack of consistent correlations with WM

If the linguistic data are characterised by lack of change, probably due to a combination of methodological, pragmatic or underlying feature specification difficulties, what may best explain the lack of expected consistent correlations between the linguistic data and WM?

The evidence of some positive correlations suggests that WM can facilitate L2 development, as a workspace for more efficient linguistic analysis - greater capacity as measured in Story Recall was implicated in greater improvement in the production of accurate questions by Time 2. However, WM seems to be used in different ways for different tasks and to different degrees at Time 1 or Time 2. The clearest pattern showed that greater WM capacity was consistently associated with slower processing in the grammaticality judgement task. In addition, only the Story Recall tasks provided any consistent pattern of correlations, which was unexpected.

The pattern of higher Story Recall scores with slower speeds is very striking across all measures at Time 1 and Time 2, and in change in speed on most measures. However, I argue that this is a task-specific strategy: even though participants were given instructions not to "try and remember any rule" but to give their initial instinctive response, it is possible that they tended to treat it as a test of grammar knowledge involving metalinguistic reflection. Evidence of the wide range of times could support this suggestion (the fastest time was just over 2 minutes, and the slowest over 15 minutes, at Time 2). The connection between slower speed and greater WM capacity echoes Kroll et al (2002) who found higher WM span correlated with slower speed on a lexical translation task. Further research using limited-time response methodology (such as moving window techniques - Juffs and Harrington 1995; Marinis 2003) would help to clarify how far the RT task as used here was tapping online or offline knowledge, and would therefore further help understand the role of WM in SLA (especially in distinguishing acquisition from use).

Taken together, the findings from the first and second studies suggest a number of methodological and conceptual problems arising in this investigation of the role of WM in SLA. One of the crucial assumptions on which this study was based was that adult WM is relatively stable between 22 and 40 years (Baddeley et al 2009). If so, the longitudinal design of these studies was plausible: that WM at time 1 and L2 scores at Time 2 could be related. However, it seems that, based on the evidence, particularly in the second study, this assumption may be questionable. Wilcoxon signed rank related-samples tests comparing scores at Time 1 and Time 2 showed that Digits Back in L1 between Time 1 and Time 2 was not significantly different ($p>.05$), but that Digits Back in L2 was significantly different ($p<.01$). Story Recall in L1 also showed significant differences between Time 1 and Time 2 ($p<.01$) and near significant differences for Story Recall in L2 ($p=.052$). Digits Back and Story Recall tests correlated with each other within language and across language at Time 1 and Time 2 ($p<.05$), but Listening Span showed no correlation between Time 1 and Time 2 ($r=.185$, $p>.05$).

There are two possible reasons for this lack of stability. It is possible that causes of variation in WM (hormonal differences, tiredness, hunger), which are evident if WM is measured at different times of day or days of the week (Matthews et al 2000: 217), may also occur when WM is measured many months apart. It is also possible that WM operates differently in the L1 and the L2 due to the effects of multicompetence, i.e. the effect on cognition of multilingualism (Cook 1997, 2002). Moreover, it could be that increased L2 proficiency could affect not only WM measured in L2 (as has been found - Service et al 2002), but also WM measured in L1. However, there is no research that I have been able to find that would specifically deal with the question of longitudinal change in WM in multilinguals, especially developing adult L2 users, so these suggestions at this stage therefore remain highly speculative. Further research into longitudinal measures of WM comparing monolinguals with bilinguals would clarify this issue.

The dominance of correlations with Story Recall rather than any of the other tasks also needs clearer understanding. The lack of any significant correlations with the Digits Back tests was not entirely unsurprising, in that such number-based tasks have not consistently been found to relate to linguistic tasks, leading some to argue that verbal and non-verbal working memory should be conceptualised as separate (Waters and Caplan 1996). However, the Digits Back task has been found in at least one study to have a relation to L2 linguistic development measured in a global score of intermediate-level general proficiency including reading, writing and oral English (Kormos and Safar 2008). Kormos and Safar (2008), as discussed in section 10 in the literature review, chapter 2) also found positive correlations between their global measure of intermediate-level English proficiency and a non-word repetition task. Such tasks test phonological loop capacity (Gathercole and Baddeley 1993), and have been specifically and robustly associated with verbal learning in first language acquisition. Kormos and Safar (2008) associate their findings with the capacity of the phonological loop to store strings of sound, which facilitates learning of chunks and which can in turn facilitate induction of simple grammatical rules (see also Ellis and Sinclair 1996).

However, most studies using non-word repetition have focused on novel vocabulary development, whereas the focus in the present study was grammatical development, especially in more advanced level complex structures, for which I wished primarily to focus on the executive elements of the WM model. Further research using phonological loop tests, such as non-word recall, may, however, provide additional data to investigate the interaction between WM and grammatical development, although Gathercole and Baddeley (1993) do not report any robust evidence in first language acquisition between phonological memory and grammatical development in children.

There was a specific methodological problem with scoring validity for the Listening Span task, which many participants reported in post-task debriefing conversations as very difficult. Two issues arise particularly, to do with scoring and to do with design. The task as originally conceived by Daneman and Carpenter (1980) involved recalling sets of final words from increasing numbers of sentences which had been either heard or read; using an absolute scoring method creates a range from 2 to a maximum of 6 or 7 words recalled. In other studies (for example by Sagarra 2000), no participant scored more than 3.5 words out of a possible maximum of 6 at either Time 1 or Time 2, and the mean was 2.3 (Time 1) or 2.6 (Time 2). The lack of L2 studies using Listening Span, as opposed to Reading Span, makes direct comparison with previous studies difficult, but if Listening Span and Reading Span are analogous (as argued by Daneman and Carpenter 1980), then possible comparisons can be made to, e.g., Sagarra (2000) who found a maximum of 6 using the original Daneman and Carpenter Reading Span test, and Walter (2004) who found a maximum of 5 (although her L2 mean score, 2.47, was similar to that found in the present study).

The lack of differentiation between span levels using an absolute scoring system (a problem also for Sagarra 2000) was the principal reason for using a ratio score based on successful recall (Conway et al 2005), as it would allow a more informative means of ranking individual participants. The disadvantage of the ratio scoring system is that it has no means of taking into account the difference between the many participants who achieved very successful scores up to a set size of 3, but then managed no more, compared to the few who had a go at sets of 4 or above but may not have managed more than 1 final word per set. Hence a ceiling effect was observed (several participants achieved above 0.90, or near perfect scores). This scoring difficulty also applied to the Digits Back tasks. Further research would allow this methodological issue to be tested and revised accordingly.

A secondary methodological problem with the Listening Span task was that, for reasons of efficiency in gathering all the data in individual test situations and avoiding over-burdening the participants, the Listening Span test was only carried out in the L2. Establishing an L1 version of the task, and repeating the test in both languages would provide greater test validity. An additional conceptual issue arises with the task's design. It was designed to be slightly different to the usual reading or listening span tasks, as it used material that would commonly be mentally visualised (processing directions, as used to get around in the real world). Hence the Listening Span as used here was intended to address concerns that have been raised in the literature (Daneman and Hannon 2007) that the Daneman and Carpenter tests were primarily task-dependent: e.g. the reading task really tested reading skills. A verbal span task that deliberately utilised Baddeley's conceptualisation of the episodic buffer as a multimodal element combining information from both visual and phonological components could have been useful to take WM research forward. However, the perceived difficulty of the task, and the lack of robust findings, mean that further research is needed in order to see whether the main difficulties with the Listening Span task are primarily methodological or conceptual.[27]

A further potential confounding methodological factor may have arisen from issues of test reliability found in all the tasks. On the oral task, a test-retest reliability analysis using Cronbach's alpha across the two measures of Time 1 and Time 2 yielded less than .70 ($p<.05$). The RT task, in a similar analysis, yielded even lower scores. For speed, Cronbach's alpha was .46. For accuracy, Cronbach's alpha was .36. For the WM tasks themselves, reliability scores across Time 1 and Time 2 ranged from .56 to .71. To provide some comparison, Waters and Caplan (1996) found test-retest reliability on two WM measures was .41 (Daneman and Carpenter's Reading Span) and 0.65 (Waters & Caplan reading/judgement span). However, reliability seems not often to be measured in SLA studies; a rare example of this is found in a meta-analysis of ten studies of WM in L2 development which showed a wide range of accepted reliability scores on WM tasks of between .21 and .86 (Watanabe and Bergsleithner 2007). Sagarra (2000) found reliability scores in her study of less than .700. It could be that the Cronbach's alpha test is not the most suitable to use on only two test repetitions, and this is an area where further careful research will help clarify established reliability norms. Nevertheless, the reliability scores in this study do not seem to be obviously deviating from those used elsewhere, and I do not conclude that task unreliability had a definite effect on the lack of significant findings.

6.5. Implications of Story Recall findings

The interaction found between linguistic measures and Story Recall is of particular interest, as the Story Recall tests were specifically designed to test the episodic buffer. The construct of the episodic buffer has not been tested in L2 before, but was seen as a test of the ability to manage complex syntax in a speaker's native language (Fry 2002, as outlined in Chapter 2, section 2.4.5). The hypothesised positive correlation between Story Recall and the measures of complex syntax (in the oral task and in the Long Movement sub-test in the Reaction Time task) was not clearly evident here.

---

[27] Baddeley (p.c.) suggests that one difficulty with the way the task was conceived may lie in evidence that visual storage is currently seen as much more costly than phonological (Baddeley et al 2009), imbalancing the storage and processing balance of the verbal and visual material. In other words, in thinking about directions, the visual element may take up disproportionate amounts of storage space and processing capacity, diminishing the capacity to handle the verbal material.

Nonetheless the consistent significant findings for Story Recall and slower processing across all structures need explaining. In addition, there is a specific implication of some role for Story Recall capacity regardless of form (explicit taught or implicit untaught forms), noted in the evidence discussed in chapter 5 at the end of section 5.4.3. This evidence showed that there were significant between-group differences on groups split by scores in Story Recall, in which the highest scoring group showed significantly higher mean scores in accuracy on subjacency-constrained items, and slower response times for those items. These data confirmed that, for L2 users, greater capacity in L1 domain-general storage and processing (episodic buffer) also resulted in significantly slower, significantly more accurate judgements on non-taught forms as well as for taught forms.

One strand of previous research which may add to the plausibility of Story Recall as a test of ease of "access" to language is the work done by Rumelhart and others in the 1970s and 1980s on story grammars and narrative-level schemas. The concept of schemas for producing language fluently is argued to operate both at phrasal and also at narrative level (as discussed in section 2.3.2 in the literature review). In addition to clause-based grammatical structure, it is argued that speakers can draw on discourse-level structures, which can be stored as prototypical narratives, aiding the construction of online events. An initial attempt was made in the 1970s (led by Rumelhart 1975) to produce a combination of syntactic and semantic "rules" that could automatically generate story grammars and story schemas. The story grammar specifies the units (e.g. attempts and outcomes) and the order of units in a story, e.g. providing the setting before episodes or actions. The story schema provides the listener with a pre-packaged construct allowing them to generate expectations in an abstract way about the information in a simple story (Sternberg and Smith 1988: 247). Although the attempts to create such rules no longer seem to be an active research field, the notion of a story schema remains useful as a heuristic for understanding how narratives are constructed, and especially the cultural and linguistic assumptions that would underpin a given schema, which could lead to misunderstandings in L2 users (Hinkel 2005: 513). The location of such schemas as items of declarative or procedural (or explicit or implicit) memory is not clarified, as far as I am aware. However, in terms of how language is constructed online, and the link between the capacity to remember the order of events and express them linguistically, the role of story schemata could explain the plausibility of Story Recall as a valid addition to WM test methodology.

In terms of the central research questions driving this research to establish a connection between WM and SLA, the statistical evidence of a relationship between Story Recall in L1 and improved performance both in the oral task and in the most complex target forms in the reaction time task (subjacency) are exciting evidence of some implicational relationship between WM and SLA. I therefore propose that Story Recall has plausible test validity for the proposed role of the episodic buffer in terms of providing a link between long term memory, medium-term storage of novel information and central executive processing, where both implicit and explicit grammatical information can be manipulated along similar lines to Levelt's Formulator (1989). Further research using this test methodology in L1 and L2 on language learners on a range of complex language forms would help take this proposition forward.

However, these data also lead to a theoretically-motivated challenge to the model of WM used here. The fundamental assumption of this study, based on Baddeley's model, was that WM would be implicated in processing or production of frequent, explicitly taught forms but not implicit forms, since Baddeley expressly connects WM with declarative memory. In his model (shown in chapter 2, section 4.2) the crystallised portion of long term memory specifically referred to is episodic memory and semantic memory (including language and visual semantics), which he sees as comprising explicit (declarative) memory (Baddeley et al 2009: 10). It was this model that prompted the present study's design to compare explicit with implicit forms in order to test the claim that WM was the potential "key to variation" in L2 acquisition (Miyake and Friedman 1998).

The assumption that WM would be implicated in analysis of explicit structures by facilitating retrieval of metalinguistic, conscious knowledge of explicitly taught forms was borne out in higher WM correlations and slower times for explicit forms in the RT task at Time 1. Nevertheless, the pattern of correlation between greater Story Recall capacity and slower speed was found equally across all forms, including those thought to be implicitly learned, at Time 2. In addition, Story Recall in L1 correlated with accuracy on subjacency, as shown above.

Using Baddeley's model, it is not clear why the implicit subjacency-constrained target structures showed similar patterns of correlation to the explicit taught structures. These data suggest not only that WM should be associated with the declarative memory system, but also that WM can interact with procedural knowledge of implicit forms, by acting as a workspace for structures combining both implicit and explicit knowledge, which is brought into conscious focus for manipulation. In sum, the assumption drawn in this study that WM will favour retrieval of explicit knowledge, and should not be associated with managing linguistic information derived from implicit or non-taught knowledge, therefore no longer holds.

The concept of WM interacting with both explicit and implicit knowledge stored in long-term memory (LTM) links back with older conceptualisations of WM as a separate system (Anderson 1983, 1993). Baddeley's model does of course see WM as separate, but it should be possible to see how far the model could be revised to incorporate information derived from implicit memory into a current "workspace".[28] Further research using a combination of WM tests with other typical tests of implicit syntactic constraints common in the SLA literature (e.g. verb raising in French, verb second word order in German, long-distance binding in Chinese and Japanese, distribution of null subjects in Italian and Spanish) could further shed light on the role of WM as workspace for both explicit and implicit knowledge. In addition, it is possible to argue (following Ullman 2005) that the projected explicit/implicit distinction does not map precisely on to the declarative memory systems. However, it may be that the dichotomies of acquisition vs. learning (Schwartz 1993) and declarative vs. procedural knowledge (Ullman 2005) which underpinned the test design for this study, are not in practice so distinct or materially separable, especially in a testing situation (Heredia and McLaughlin 1992; Herschensohn 1999; Paradis 2009).

It may also be concluded that Baddeley's WM model used as the basis for this study, and the tasks derived to test that model are not very amenable to L2 studies, particularly studies of syntactic competence. I have made an argument for further investigation of the role of Story Recall as a possible test methodology that will further our

[28] The separation of WM from both types of LTM, but interacting with both, is reiterated in more current research by Ullman, p.c.)

understanding of the interaction of memory and SLA. However, in terms of more standard WM tests based on Baddeley's model, this study seems to add to the growing list of studies which have looked for the predictive power of WM to aid development in L2 but have not been able to find it, such as Sagarra (2000) who looked at semi-longitudinal acquisition of Spanish, and Juffs (2004) who looked at processing of wh-extraction.

The implication of WM capacity affecting both implicit and explicit structures may be argued to fit better with MOGUL's adoption of Cowan's (1999, 2005) conceptualisation of the role of working memory as the "currently activated" portion of long-term memory (also as discussed by Jackendoff 2002). As Truscott and Sharwood Smith conceive it within MOGUL (2004, ms), WM operates modularly as the currently active portion of each of the different elements of linguistic knowledge (CS, PS and SS), and may be conceptually distinct from working memory used for non-linguistic information in the perceptual sensory input (counter to Baddeley's domain-interactive model). Resources available to the L2 user in developing their linguistic knowledge could derive from input computed from explicit information in the input (positive evidence of how to construct simple and long-distance questions), combined with implicit knowledge shaped by innate universal constraints (such as those operating on wh-movement). The evidence of WM implication in slower response times, for example, could mean that participants with greater working memory spend longer trying to "work out" an appropriate response using all the resources available to them. By implication, those with lower capacity would have less storage and processing resources and would leap to a judgement in less time.

The discussion of the data found in this study must then be widened beyond the hypotheses tested here. If WM does not appear to play a material or implicational role in aiding L2 development, except in two measurable cases (oral accuracy and development in accuracy on implicit forms), what other factors could be argued to have a measurable effect?

6.6. Potential confounding factors beyond the original scope of the study

One potential confounding factor could arise from the question of how participants were responding to the input (Piske and Young-Scholten 2009), especially in view of the unexpected lack of material change in the linguistic knowledge displayed. As noted above, and discussed in more detail in Chapter 2, a number of studies have also failed to find evidence of significant improvement in morphosyntactic development over periods of immersion (Collentine and Freed 2004; Rothman and Iverson 2007; Sunderman and Kroll 2009).

In order to control for potential differences in type and amount of exposure to input, three variables had been measured prior to the start of the study - age of learning (AOL), length of learning (LOL) and additional exposure during school. As noted in Chapter 4, which details the revised methodology and biodata for the second study, these variables had not shown any significant effect on participants' IELTS scores ($p>.05$). AOL had no significant correlations (no r coefficient below -0.2, $p=.35$ or above), nor LOL (no r coefficient above 0.2, $p=.4$ or above - see Appendix D).

The group had therefore been deemed to be as homogenous as possible prior to immersion. The amount of exposure during immersion was assumed to be constant during the first part of participants' immersion (since attendance in class was mandatory, and the number of taught classes is relatively homogenous for Masters' programmes, at around ten hours per week), so I investigated differences in exposure towards the end of their immersion period, when they no longer attended class, but were preparing their Masters' dissertations. Diary data (self-reported) were collected across a week to calculate a daily average of any kind of English language use (reading, writing, listening or speaking, and in academic or non-academic settings, including with native or non-native speakers of English). Just over half the participants returned the diaries (19 out of the original pool of 32). As could be expected, individuals varied from a minimum of 3.29 hours per day to 14.71 hours; the mean daily average was 7.77 hours per day (with usage heavily skewed towards reading over other types of activity). There were no correlations found between self-reported hours of exposure and any of the oral or RT scores at Time 2, or improvements between Time 1 and Time 2 ($p=.3$ or above; see Appendix D), although this may have been confounded by the wide variation in the diary data, and the potential lack of validity and reliability in self-reported data (Mackey and Gass 2005). Nonetheless I conclude that individual differences in exposure did not play a critical role in variation in development during immersion.

It may be assumed that immersion does not guarantee exposure to the kind of rich input required to trigger acquisition (Flege and Liu 2001). As noted in the literature review, Flege and Liu (2001) investigated a number of measures of targetlike phonological and grammatical competence comparing students to non-students, distinguished as short-stay or long-stay. They found that long-stay students scored significantly more highly than other groups, and assumed this was due to high levels of engagement with native speakers using rich complex language, typical of academic environments. However, in my study, all the participants were students, so the type of rich native input was assumed to be constant, and yet there was still no material change in accuracy.

It is possible that longer immersion would have had a greater linear effect on improvements. To check this, the RT test used in this study was also used on three other Chinese L1 students resident for more than three years in the UK. The data collection was done cross-sectionally, so a direct comparison with the longitudinal patterns of development shown by the group in the present study is not possible, but, as an approximate indicator, the comparison group's scores for RT time (total, and for grammatical and ungrammatical items) were compared to the participants from the second study scores at Time 2, presented in the table below.

Table 63: Comparison of second study participants to longer-stay students

RT scores		total	grammatical	ungrammatical
Comparison group of long-stay students *(N=3)*	Mean RT times	332.89	169.05	163.84
	Mean accuracy	53.67	21.67	32
Second study participants at Time 2*(N=32)*	Mean RT times	432.81	208.14	224.68
	Mean accuracy	39.69	23.59	16.09

Scores suggest that both RT speed and accuracy did improve, although not greatly so for grammatical items, suggesting that greater length of exposure was potentially a confounding factor. Future research using a longer time scale would help to clarify this possible issue.

It is also possible that the assumption that students would receive rich input with plenty of evidence of the target structures was not justified. The assumption drawn from the literature was that questions are frequent and salient in academic language (e.g. Hinkel 2003), and was also informed by my own background of ten years' experience of teaching English to international students across 4 UK universities, as well as familiarity with the academic discourse presented in the two universities used in this study, in which question-answer structures are common in almost all academic encounters, as well as informal social encounters which the international students are encouraged to join in. However, this issue could be empirically tested in future research.

To sum up, it remains an open question as to why the participants in this study showed so little linguistic development, in that only speed of reaction times and oral fluency significantly improved but accurate oral production or grammatical judgments showed no significant change. It has been discussed that the data show difficulties both in acquisition (of appropriate L2 feature strengths) and in processing; there is some evidence of greater automaticity and less control, but reliance on conscious monitoring and optionality between targetlikeness and nontargetlikeness remained clear even after immersion.

Research both into the different effects of quality vs. quantity of input (see, e.g. Piske and Young-Scholten 2009), and also into differences between explicit vs. implicit processing (e.g. N. Ellis 1994; VanPatten 1996; Schmidt 2001; Hulstijn 2005) clearly suggests that there is more to input than the generative paradigm assumes (as discussed in more detail in Chapter 2). In addition, clarifying exactly what input is and how it drives acquisition remains an ongoing question, which Truscott and Sharwood Smith's work on MOGUL goes some way to address. However, even the most recent studies acknowledge that variability in L2 remains difficult to explain, and that immersion per se does not appear to aid transition through the stages of development, or guarantee what the end-state of ultimate attainment can be, as discussed in Chapter 2 (see e.g. Collentine and Freed 2004; Birdsong 2006; Rothman and Iverson 2007; Lardiere 2007). The role of WM also remains unclear, although in one recent study (Sunderman and Kroll 2009), it appears that the effect of WM on greater development during immersion may depend on a kind of internal threshold of WM capacity - above which immersion did facilitate improvement, but below which added exposure through immersion did not make a significant difference.

It is possible to speculate that the kind of traditional grammar-instruction method experienced by the participants in this study, and lack of access to input from native speakers of English during their educational exposure to English prior to arrival in the UK, may lead to entrenchment or fossilisation of certain forms, such as copula question chunks, non-targetlike marking of number, or verbal inflection attachment to lexical verbs. This would make targetlike verb raising, and distinguishing matrix from embedded questions, very difficult. Han (2003, 2006) argues that forms that cross the syntactic/discourse interface, such as question forms, may be vulnerable to fossilisation. It is also noted that the target forms investigated here all involve projecting CP, which has been identified as potentially vulnerable (Platzack 2001 – see discussion in Chapter 2, sections 2.2.1, 2.12). In this case, the participants may, to some extent, have fossilised before arrival; it could well take much longer than a year to overcome the strength of entrenchment, particularly for forms within a domain identified as subject to variability. Investigation of how question forms are taught and used within the Chinese school learning environment (i.e. before immersion in the L2 environment) would provide some indication of the connection between input and acquisition prior to immersion.

It was also noted in a study of international students using English L2 (Chapter 2, section 2.6) that, descriptively speaking, such non-native students often have ongoing difficulties in academic contexts requiring structures similar to those discussed here, including differentiation of tenses, embedded clauses and subordination (Hinkel 2003). It could be speculated that the kind of passive exposure most prevalent in a taught postgraduate programme may not provide the kind of input to trigger development on such structures. It may also be that an optimally rich input environment should also involve a greater role for oral interaction (Long 1990, 1996; Gass 1997; Mackey 1999; Mackey et al 2002). However, it is suggested that written input does not necessarily promote slower or less accurate linguistic knowledge than oral input (Cook 2008). In addition, in terms of different types of written discourse, or genre (Swales 1990), the kind of complex language used in the academic genre should have provided plenty of examples of the kind of complex forms targeted in this study.

Detailed analysis of the issues of input, interaction and awareness lay beyond the scope of this project, although some of the research paradigms were incorporated to some degree, i.e. of the assumption of increased frequency of the target structures in immersion input compared to instructed input, and of the potential role for awareness and control in the model of WM used here. This study's findings that WM seems to be implicated in both explicit and implicit language knowledge development means that further research based on clearer definitions of consciousness and awareness is required. Additionally, detailed empirical evidence that the structures investigated here are widely salient in the kind of academic input to which the participants were exposed is also required. Finally, further investigation into the role of different input contexts, and the importance of individual differences in responding to input, is also needed to clarify these issues further.

Chapter 7: Conclusion

To conclude, it seems that, counter to Schwartz (1993), exposure to primary linguistic data may not be enough to explain development, and that in order to develop a more explanatory theory of transition in SLA, it is necessary to look more closely into how universal implicit constraints may interact with other factors. Such a belief prompted this cross-disciplinary study, in order to evaluate how effective combination or competition theories can be in explaining the data (such as MOGUL, amongst others, as discussed in the literature review). The findings here were argued to be compatible, at least in part, with the MOGUL model of implicit constraints and features operating on the input at different levels of activation in response to L1 effects and individual differences in memory and processing, evident in optionality, hesitancy and variability. This study is therefore intended to contribute to driving a greater understanding of how at least some of the different factors discussed above combine in the complex process that is adult acquisition of a second language.

As highlighted at the end of chapter 6, the issues covered in this study about the nature of L2 acquisition, the nature of L2 linguistic knowledge and processing of that knowledge in real time remain opaque. The original premise of this study that working memory is involved in language development was not straightforwardly supported, although greater working memory capacity was found to correlate with greater improvement in targetlike oral question formation. However, perhaps the most surprising result from this study, rather than the complex nature of working memory, was the lack of material change in linguistic knowledge, as measured in oral and reaction time question tasks, during a year of intense immersion in the L2 environment. While this result seems limited, it does point to a real difficulty for students, whose linguistic ability is challenged by the complex language demands of postgraduate study abroad. The ability to ask questions accurately and manage complex grammar is fundamental to being an effective student (Hinkel 2003). The evidence from the participants in this study showed just how much of a hurdle some of them face in achieving sufficient proficiency in such key grammatical structures.

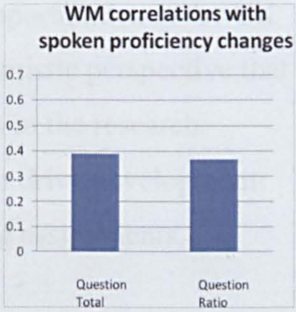
As a clear illustration summarising the complex and wide range of issues studied and mixed data results, the principal findings from the second study are shown below in figure form.

Research Predictions:

Greater WM correlates with improvements in:

a: speech (asking accurate questions)?

Yes - Significant (p<.05)



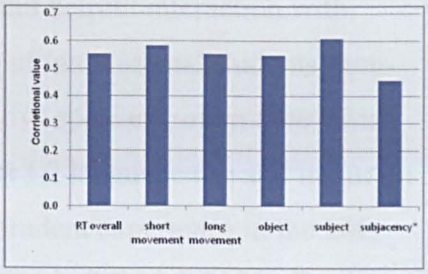
b: accuracy and speed in RT processing (judging question forms)?

No- not significant for accuracy (p>.05)

Unexpected findings:

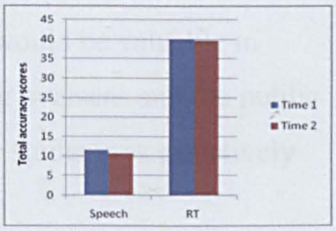
1. WM correlates with slower RT processing, not faster as predicted, for both explicit and implicit types.

Significant (p<.05)



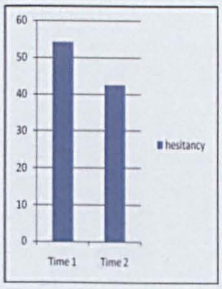
2. Limited effect for immersion

a: no material improvement in oral or RT accuracy



b: only improvements found on oral fluency and RT speed (less hesitancy and shorter response times)

Significant (p<.05)



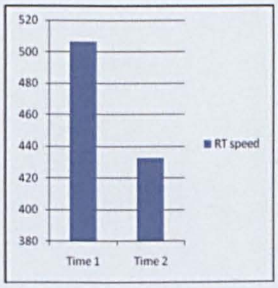


Figure 21: Summary of second study findings

I therefore conclude with some speculation about the nature of input required to trigger language change. It is assumed from a lay perspective that studying for a Masters degree in the UK provides an opportunity for international students to "perfect" their knowledge of the English language - a claim found on Graduate Prospects, a UK-government supported website, marketing international study opportunities in the UK to overseas students. However, it is possible to argue from a linguistic perspective that the mainly academic type of exposure offered to the participants in the research presented here is too divergent from optimal linguistic contact to drive development: that input derived primarily from lectures, and reading or writing assignments, arguably, do not provide the kind of input that triggers change.

Concerns have been noted about the difficulties many international students face in an immersion environment in participating in appropriate and helpful interaction with native speakers; yet it is often the case that the majority of international students' non-classroom time is spent in the company of other non-native speakers, often their own L1 compatriots, thus further reducing the opportunity for L2 language use and input. In a recent government-sponsored survey of international student experience in the UK, only 15% of Chinese students reported having British friends (Merrick 2004). Further detailed research on the type of input and interaction available to Masters students, controlling more closely for the quality and quantity of input, would be valuable in addressing these concerns, both for international students going overseas and the public policy makers and universities who should be supporting those students as effectively as possible.

REFERENCES

Acklam, R. (1996). *First Certificate Gold Coursebook*. Harlow, Longman.

Adams, A.-M. and Gathercole, S. (1996). Phonological working memory and spoken language development in children. *Quarterly Journal of Experimental Psychology, 49A*: 216-233.

Adams, A.-M., and Gathercole, S. (2000). Limitations in working memory: implications for language development. *International Journal of Language and Communication Disorders* 35: 95-116.

Adger, D. (2003). *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.

Alloway, T. (2007). The *Automated Working Memory Assessment*. London: Psychological Corporation.

Alloway, T. (2009). Working Memory: Is it the New IQ? Presentation, Newcastle University CRiLLS seminar series.

Alloway, T. and Gathercole. S. (2005). Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology* 17(2): 207-220.

Anderson, J. (1983). *The architecture of cognition*. Mahwah, NJ: Lawrence Erlbaum.

Anderson, J. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J., Bothell, D., Byrne, M., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychological Review 111* (4): 1036-1060.

Andrade, J. (ed.) (2001). *Working memory in perspective*. Hove: Psychology Press.

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders* 36 (3): 233-240.

Atkinson, M. (1996). Now, hang on a minute: some reflections on emerging orthodoxies. In H. Clahsen (ed.) *Generative perspectives on language acquisition*. Amsterdam/Philadelphia, John Benjamins. 451-487.

Atkinson, R. and Schiffrin, R. (1968). Human memory: a proposed system and its control processes. In K. Spence and J. Spence (eds.) *The psychology of learning and motivation*. London: Academic Press. 2.

Aoun, J. and Li, Y-H. (2003). *Essays on the representational and derivational nature of grammar*. Cambridge, Mass: MIT Press.

Baars, B. (1997). *In the theater of consciousness: the workspace of the mind*. New York: Oxford University Press.

Baddeley, A. (1986). *Working Memory*. Oxford: Clarendon Press.

Baddeley, A. (1999). *Essentials of human memory*. Hove: Psychology Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences* 4 (11): 417-423.

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communications Disorders* 36: 189-208.

Baddeley, A. (2007). *Working Memory, thought and action*. Oxford: Oxford University Press.

Baddeley, A., Eysenck, M. and Anderson, M. (2009). *Memory*. Hove: Psychology Press.

Baddeley, A., Gathercole, S. and Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review* 105 (1): 158-173.

Baddeley, A. and Hitch. G. (1974). Working memory. In G. Bower (ed.) *The psychology of learning and motivation.* New York, Academic Press: 47-89.

Baddeley, A. and Logie. R. (1999). Working memory: the multiple-component model. In A. Miyake and P. Shah (eds.) *Models of working memory*. Cambridge: Cambridge University Press.

Baddeley, A., Papagno, D. and Vallar. G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*. 27: 586-595.

Bailey, N., Madden, C. and Krashen, S. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning* 24: 235-243.

Bates, E. and MacWhinney. B. (1989). Functionalism and the competition model. In E. Bates and B. MacWhinney (eds.) *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press.

Bayley, R. (1994). Interlanguage variation and the quantitative paradigm: past tense marking in Chinese English. In S. Gass, E. Cohen, and E. Tarone (eds.) *Research Methodology in Second Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Beck, M.-L. (ed.) (1998). *Morphology and its interfaces in second language knowledge.* Amsterdam/Philadelphia, John Benjamins.

Bertolo, S. (ed.) (2001). *Language acquisition and learnability*. Port Chester, NY: Cambridge University Press.

Berwick, R. (1985). *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.

Bhatt, R. and Hancin-Bhatt, B. (2002). Structural minimality, CP and initial state in second language acquisition. *Second Language Research* 18 (4), 348-392.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition* 16 (2): 157–168.

Bialystok, E. (2002). Cognitive processes of L2 users. In V. Cook (ed.) *Portraits of the L2 User*. Clevedon: Multilingual Matters. 145-166.

Birdsong, D. (1989). *Metalinguistic performance and interlanguage competence*. New York, Springer.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language* 68: 706-755.

Birdsong, D. (2006). Nativelikeness and non-nativelikeness in L2A research. *International Review of Applied Linguistics* 43: 319- 328.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis* 20 (1-2): 3-49.

Bley-Vroman, R. (2009). The evolving context of the Fundamental Difference Hypothesis. *Studies in Second Language Acquisition* 31 (2): 175-198.

Bley-Vroman, R., Felix, S. and Ioup, G. (1988). The accessibility of Universal Grammar in adult language learning. *Second Language Research* 4(1), 1–32.

Bloom, P. (1994). *Language acquisition: Core readings*. Cambridge, MA: MIT Press.

Boeckx, C. (2003). *Islands and Chains*. Amsterdam/Philadelphia: John Benjamins.

Bond, M., (ed.) (1996). *Handbook of Chinese psychology*. Hong Kong: Oxford University Press.

Bresnan, J. (ed.) (1982). *The mental representation of grammatical relations*. Cambridge, Mass: MIT Press.

Bresnan, J. (2001). *Lexical Functional Syntax*. Oxford, Blackwell.

British National Corpus available at http://www.natcorp.ox.ac.uk/
Last accessed 20/12/07

Brooks, L. and Vokey, J. (1991). Abstract analogies and abstracted grammars. *Journal of Experimental Psychology: General* 120: 316-323.

Brown, G. and Hulme, C. (1992). Cognitive processing and second language processing: the role of short term memory. In R. Harris (ed.) *Cognitive processing in bilinguals.* Amsterdam: Elsevier. 105-121

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form.* Amsterdam: John Benjamins.

Caplan, D. and Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences* 22 (1): 77-94.

Carroll, J. and Sapon, S. (1959). Modern Language Aptitude Test. New York, The Psychological Corporation/Harcourt Brace Jovanovitch.

Carroll, S. (2001). *Input and evidence*. Amsterdam: John Benjamins.

Cheng, L. (1997). *On the Typology of Wh-Questions*. Garland Publishing: New York.

Cheng, L. (2009). "*Wh*-in-situ, from the 1980s to now". *Language and Linguistics Compass* 3 (3): 767–791

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology* 32 (5): 867-873.

Cheung, H., Kemper, S. and Leung, E. (2000). A phonological account for the cross-language variation in working memory processing. *Psychological Record* 50 (2): 373.

Chincotta, D. and Underwood, G. (1997). Digit span and articulatory suppression: a cross-linguistic comparison. *European Journal of Cognitive Psychology* 9: 89-96.

Chomsky, N. (1980). *Rules and Representations.* Oxford: Blackwell.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Barriers*. Cambridge, Mass: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass: MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry* 36 (1): 1-22.

Christoffels, I. (2006). Listening while talking: The retention of prose under articulatory suppression in relation to simultaneous interpreting. *European Journal of Cognitive Psychology* 18 (2): 206-220.

Clahsen, H. (ed) (1996). *Generative perspectives on language acquisition*. Amsterdam/Philadelphia, John Benjamins.

Clahsen, H. and Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics* 27: 3-42.

Clahsen, H. and Muysken, P. (1986). The availability of universal grammar to adult and child learners - a study of the acquisition of German word order. Second *Language Research* 2 (2): 93-119.

Clark, E. (2003). *First Language Acquisition*. Cambridge: Cambridge University Press.

Clifton, C., Frazier, L., and Rayner, K. (1994). *Perspectives on sentence processing.* Hillsdale: Lawrence Erlbaum.

Cohen, N., Eichenbaum, H., Deacedo, B. and Corkin, S. (1985). Different memory systems underlying acquisition of procedural and declarative knowledge. *Annals of the New York Academy of Sciences* 444, 54-71

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition* 26 (2): 201-225.

Collentine, J. and Freed, B. (2004). Introduction: learning context and its effects on second language acquisition. *Studies in Second Language Acquisition* 26 (2): 153-172.

Conway, A., Kane, M., Bunting, M., Hambrick, Z., Wilhelm, O. and Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review* 12: 769 - 786.

Cook, V. (1991). The poverty-of-the-stimulus argument and multi-competence. *Second Language Research* 7 (2): 103-117.

Cook, V. (1997). The consequences of bilingualism for cognitive processing. In A. de Groot and J. Kroll (eds.) *Tutorials in bilingualism: psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum. 279-300.

Cook, V. (ed.) (2002). *Portraits of the L2 user*. Clevedon: Multilingual Matters .

Cook, V. (2008). *Second language learning and language teaching*. London: Hodder Education.

Cook, V. and Newson, M. (2007). *Chomsky's Universal Grammar*. Oxford: Blackwell.

Coppetiers, R. (1987), Competence differences between native and near-native speakers. *Language* 63 (3): 545-573.

Coughlan, A. and Hollows. S. (1985). The Adult Memory and Information Processing Battery (AMIPB) test manual. Leeds: Coughlan.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Mikaye and P. Shah (eds.) *Models of working memory*. Cambridge: Cambridge University Press. 62-101.

Cowan, N. (2005). *Working memory capacity*. Hove : Psychology Press.

Cowan, N., Elliott, E., Saults, J., Morey, C., Mattox, S., Hismjatullina, A. and Conway, A. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology* 51 (1): 42-100.

Craats, I., Kurvers, J. and Young-Scholten, M. (eds.) (2006). *Low-educated second language and literacy acquisition: proceedings of the inaugural LESLLA symposium*. Netherlands Graduate School of Linguistics: LOT Publications.

Craik, F. (1979). Levels of processing: Overview and closing comments. In L. Cermak and F. Craik (eds.) *Levels of processing in human memory*. Hillsdale, NJ: Lawrence Erlbaum. 447-461.

Craik, F. (2002). Levels of processing: Past, present ... and future? *Memory* 10 (5-6): 305-318.

Craik, F. and Lockhart. R. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11: 671-684.

Craik, F. and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, General* 104: 268-294.

Crain, S. and Lillo-Martin. D. (1999). *An introduction to linguistic theory and language acquisition.* Malden, MA: Blackwell.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J., Aso, T., Urayama, S., Fukuyama, H., Stockton, K., Usui, K., Green, D. and Prince, C. (2006). Language Control in the Bilingual Brain. *Science* 312: 1537-1540.

Culicover, P. and Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research* 20 (6): 445-464.

Daneman, M. and Carpenter. P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19: 450-466.

Daneman, M. and Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language* 25: 1-18.

Daneman, M. and Hannon, B. (2007). What do working memory span tasks like reading span really measure? In N. Osaka, R. Logie, and M. d'Esposito (eds.) *The cognitive neuroscience of working memory*. Oxford: Oxford University Press. 21-42.

De Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics* 13: 1-24.

De Bot, K. (2008). The imaging of what in the multilingual mind? *Second Language Research* 24 (1): 111-133.

De Groot, A. and Kroll, J. (eds.) *Tutorials in bilingualism: psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum.

Dechert, H. (1980). Pauses and intonation as indicators of verbal planning in second language speech productions. In H. Dechert and M. Raupach (eds.) *Temporal variables in speech*. The Hague: Mouton. 271-285.

Dechert, H. and Raupach, M. (1987). *Psycholinguistic models of production*. Norwood, NJ: Ablex.

DeKeyser, R. (1995). Learning second language grammar rules. *Studies in Second Language Acquisition* 17: 379-410.

DeKeyser, R. (1997). Beyond explicit rule learning: automatising second language morphosyntax. *Studies in Second Language Acquisition* 19: 195-221.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22: 499-533.

DeKeyser, R. (2001). Automaticity and automatisation. In P. Robinson (ed.) *Cognition and second language instruction*. Cambridge: Cambridge University Press. 125-151.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty and M. Long (eds.) *Handbook of second language acquisition*. Oxford: Blackwell. 313-348.

Dewaele, J-M. and Furnham, A. (1999) Extraversion: the unloved variable in applied linguistic research, *Language Learning* 49 (3): 509-544.

Dienes, Z. and Perner J. (1999). A theory of implicit and explicit knowledge. *Behavioural and Brain Sciences* 22: 735–808.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: Cambridge University Press.

Dörnyei, Z. (2003). *Questionnaires in second language research*. Mahwah, NJ: Lawrence Erlbaum.

Dörnyei, Z. (2005). *The psychology of the language learner*. Mahwah, NJ: Lawrence Erlbaum.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (ed.) *Cognition and second language instruction*. Cambridge: Cambridge University Press. 206-257.

Doughty, C. and Williams, J. (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.

Dulay, H. and Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning* 24: 37-53.

Ellis, N. (1994). *Implicit and explicit learning of languages*. London, Academic Press.

Ellis, N. (2001). Memory for language. In P. Robinson (ed.) *Cognition and second language instruction*. Cambridge: Cambridge University Press. 33-68.

Ellis, N. (2005). At the interface: dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition* 27: 305-352.

Ellis, N. and Laporte, N. (1997). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition In A. de Groot and J. Kroll (eds.) *Tutorials in bilingualism: psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum.

Ellis, N. and Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: putting language in good order. *Quarterly Journal of Experimental Psychology* A 49 (1): 234-250.

Ellis, R. (1990). *Instructed second language acquisition*. Oxford: Blackwell.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition* 27 (2): 141-172.

Ellis, R. (2008). Investigating grammatical difficulty in second language learning. *International Journal of Applied Linguistics* 18 (1): 4-22.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J. and Reinders H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching.* Clevedon: Multilingual Matters.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking innateness: a connectionist perspective on development*. Cambridge, Mass: MIT Press.

Emerson, M., Miyake, A. and Rettinger, D. (1999). Individual differences in integrating and coordinating multiple sources of information. *Journal of Experimental Psychology: Learning, Memory and Cognition* 25 (5): 1300-1321.

Emonds, J. (2000). *Lexicon and grammar: the English syntacticon*. Berlin: Mouton de Gruyter.

Engle, R., Tuholski, S., Laughlin, J. and Conway, A. (1999). Working memory, short-term memory, and general fluid intelligence. *Journal of Experimental Psychology: general* 128 (3): 309-331.

Ericsson, K. and Kintsch. W. (1995). Long-term working memory. *Psychological Review* 102: 211-245.

Eubank, L. (1993/1994). On the transfer of parametric values in L2 development. *Language Acquisition* 3 (3): 183-208.

Fabbro, F. (ed.) (2002). *Advances in the neurolinguistics of bilingualism*. Udine: Forum.

Fehringer, C. and C. Fry (2007). Frills, furbelows and activated memory: syntactically optional elements in the spontaneous language production of bilingual speakers. *Language Sciences* 29 (4): 497-574.

Felix, S. (1985). More evidence on competing cognitive systems. *Second Language Research* 1: 47-72

Felser, C. and Roberts, L. (2007). Processing *wh*-dependencies in a second language: A cross-modal priming study. *Second Language Research* 23: 9-36.

Fernandez, E. (1999). Processing strategies in second language acquisition: some preliminary results. In E. Klein (ed.) *The development of second language grammars*. Amsterdam/Philadelphia: John Benjamins. 217-240.

Fillmore, C., Kay. P. and O'Connor, M. (1988). Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language* 64: 501-538.

Flege, J. and Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23: 527-552.

Flynn, S. and Lust, B. (2002). A minimalist approach to L2 solves a dilemma of UG. In V. Cook (ed.) *Portraits of the L2 user*. Clevedon: Multilingual Matters. 93-120.

Fodor, J. (1983). *The modularity of mind*. Cambridge, Mass: MIT Press.

Fodor, J. (1999). Learnability theory. In E. Klein and G. Martohardjono (eds.) *The development of second language grammars: a generative approach*. Amsterdam/Philadelphia: John Benjamins. 363-406.

Fodor, J. and Sakas, W. (2005) The subset principle in syntax. *Journal of Linguistics* 4: 513-569.

Fortkamp, M. (1999). Working memory capacity and aspects of L2 speech production. *Communication and Cognition* 32 (3-4): 259-295.

Forster, K. and Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35 (1): 116-124.

Frazier, L. and Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4 (2): 93-126.

Frazier, L. and de Villiers, J. (eds.) (1990). *Language processing and language acquisition*. Dordrecht: Kluwer.

Freed, B. (ed.) (1995). *Second language acquisition in a study abroad context*. Amsterdam/Philadelphia: John Benjamins.

Freed, B., Segalowitz, N. and Dewey, D. (2004) Context of learning and second language fluency in French. *Studies in Second Language Acquisition* 26 (2): 275-301.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6: 78-84.

Fry, C. (2002). Language complexity, working memory and social intelligence. Newcastle University (unpublished doctoral dissertation).

Gardner, R. (1985). *Social psychology and second language learning: the role of attitudes and motivation*. London: Arnold.

Gass, S. (1997). *Input, interaction and the second language learner.* Mahwah, NJ: Lawrence Erlbaum.

Gass, S., Mackey, A. and Pica, T. (1998). The role of input and interaction in second language acquisition. *Modern Language Journal* 82: 299–307.

Gathercole, S. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences* 3: 410-419.

Gathercole, S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics* 27: 513–543

Gathercole, S. and Baddeley, A. (1993). *Working memory and language.* Hillsdale, NJ: Lawrence Erlbaum.

Gathercole, S. and Thorn, A. (1998). Phonological short-term memory and foreign language learning. In A. Healy and L. Bourne (eds.) *Foreign language learning.* Mahwah, NJ: Lawrence Erlbaum. 141-160.

Gathercole, S., Hitch, S., Service, E. and Martin, A. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology* 33: 966-979.

Gathercole, S., Frankish, C., Pickering, S. and Peaker, S. (1999). Phonotactic influences on short term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25: 84–95.

Gathercole, S., Pickering, S., Hall, M. and Peaker, S. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology* 54A: 1–30.

Gavruseva, L. and Lardiere, D. (1996). The emergence of extended phrase structure in child L1 acquisition. In A. Stringfellow, D. Cahana-Amitay, E. Hughes and A. Zukowski (eds.) *Proceeding of the 20th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press: 225-236.

Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education.* Rowley, MA: Newbury House.

Gernsbacher, M. (1994). *Handbook of psycholinguistics.* San Diego: Academic Press.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68: 1-76.

Gibson, E. and Wexler, K. (1994). Triggers. *Linguistic Inquiry* 25 (3): 407-54.

Ginzburg J. and Sag, I. (2000). *Interrogative investigations: The form, meaning and use of English interrogative constructions*. Stanford: CSLI Publications.

Goad, H. and White, L. (2006). Ultimate attainment in interlanguage grammars: a prosodic approach. *Second Language Research* 22 (3): 243-268.

Goad, H., White, L. and Steele, J. (2003). Missing inflection in L2 acquisition: defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics* 48: 243-263.

Goldberg, A. (1994). *Constructions*. Chicago: University of Chicago Press.

Gombert, J. (1992). *Metalinguistic development*. Chicago: University of Chicago Press.

Goodluck, H. (1991). *Language acquisition: a linguistic introduction*. Oxford: Blackwell.

Graduate Prospects Ltd (2009) www.prospects.ac.uk. Available at http://www.prospects.ac.uk/cms/ShowPage/Home_page/International_students/Step_by_step_guide/Why_study_in_the_UK_/p!empkXdi. (Accessed 30 March 2009).

Green, D. (2003). Neural basis of lexicon and grammar in L2 acquisition: the convergence hypothesis. In R. van Hout, A. Hulk, F. Kuiken and R. Towell (eds) *The lexicon-syntax interface in second language acquisition.* Amsterdam/Philadelphia, John Benjamins: 197-218.

Greenberg, J. (ed.) (1963). *Universals of Language.* London: MIT Press.

Gregg, K. (1996). The logical and developmental problems of second language acquisition. In W. Ritchie and T. Bhatia. (eds.) *Handbook of second language acquisition*. San Diego: Academic Press. 50-84.

Gregg, K. (2001). Learnability and second language acquisition theory. In P. Robinson (ed.) *Cognition and second language instruction*. Cambridge: Cambridge University Press. 152-182.

Grosjean, F. (1982). *Life with Two Languages: An Introduction to Bilingualism*. Cambridge, MA: Harvard University Press.

Grosjean, F. (2008). *Studying Bilinguals*. Oxford: Oxford University Press.

Gu, Y. (2003). Fine brush and freehand: the vocabulary-learning art of two successful Chinese EFL learners. TESOL Quarterly 37(1): 73-10.

Guasti, M. (2003). *Language acquisition: the growth of grammar*. Cambridge, MA: MIT Press.

Gupta, P. and Cohen, N. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological Review* 109: 401-448.

Haegeman, L. (1994). *Introduction to Government and Binding Theory*. Oxford: Blackwell.

Halliday, M. (1985). *An introduction to functional grammar*. London/Baltimore: Arnold.

Hahne, A. (2001). What's different in second-language processing. Evidence from event-related brain potentials. *Journal of Psycholinguistic Research* 30 (3): 251-266.

Han, Z-H. (2004). *Fossilisation in adult second langage acquisition*. Clevedon: Multilingual Matters.

Han, Z.-H. (2006) Fossilization: can grammaticality judgment be a reliable source of evidence?, In Z.-H. Han and T. Odlin (eds.) *Studies of fossilization in second language acquisition.* Clevedon: Multilingual Matters. 56-82.

Harley, B. and Swain, M. (1984). The interlanguage of immersion students and its implication for second language teaching. In A. Davies, C. Criper and A. Howatt (eds.) *Interlanguage*. Edinburgh: Edinburgh University Press. 291-311.

Harley, B. and Wang. W. (1997). The Critical Period Hypothesis: where are we now? In A. de Groot and J. Kroll (eds.) *Tutorials in bilingualism: psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum. 19-52.

Harrington, M. (2001). Sentence processing. In P. Robinson (ed.) *Cognition and second language instruction.* Cambridge: Cambridge University Press. 91-124.

Harrington, M. and Sawyer, M. (1992). L2 working memory and L2 reading skill. *Studies in Second Language Acquisition* 14 (1): 25-38.

Hauser, M., Chomsky, N. and Fitch, T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298: 1569-1579.

Hawkins, J. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. (2004). *Efficiency and complexity in grammars.* Oxford: Oxford University Press.

Hawkins, R. (2001). *Second language syntax*. Oxford, Blackwell.

Hawkins, R. (2005). Revisiting wh-movement: the availability of an uninterpretable [wh] feature in interlanguage grammars. In L. Dekydtspotter, R. Sprouse and A. Liljestrand (eds.) *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference*. Somerville, MA: Cascadilla. 124-137.

Hawkins, R. (2008). The nativist perspective on second language acquisition. *Lingua* 118 (4): 465-477.

Hawkins, R. (2008). Can innate linguistic knowledge be eliminated from theories of SLA? *Lingua* 118 (4): 613–619.

Hawkins, R. and Casillas. G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua* 118 (4): 595-612.

Hawkins, R. and Chan, C. (1997). The partial availability of Universal Grammar in second language acquisition. *Second Language Research* 13: 187-226.

Hawkins, R. and Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, A. Hulk, F. Kuiken and R. Towell (eds.) *The lexicon-syntax interface in second language acquisition*. Amsterdam: John Benjamins.

Haznedar, B. (1997). L2 acquisition by a Turkish-speaking child. University of Durham (unpublished doctoral thesis).

Haznedar, B and Schwartz, B. (1997). Are there Optional Infinitives in Child L2 Acquisition. In *Proceedings of 21st Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. 257-268.

Henry, A. (1995). *Belfast English and Standard English. Dialect Variation and Parameter Setting*. Oxford: University Press.

Heredia, R. and McLaughlin, B. (1992). Bilingual memory revisited. In R. Harris (ed.) *Cognitive processing in bilinguals.* Amsterdam: Elsevier. 91-104.

Herschensohn, J. (1999). *The second time around: minimalism and L2 acquisition*. Amsterdam: John Benjamins

Herschensohn, J. (2007). *Language development and age*. Cambridge: Cambridge University Press.

Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly* 37 (2), 275-301.

Hinkel, E. (2005). *Handbook of research in second language teaching and learning*. Mahwah NJ: Lawrence Erlbaum.

Hofmeister, P. and Sag, I. (2009). Can processing explain Subjacency effects? Presentation, Formal vs. Processing accounts of wh-movement. University of York, UK.

Holland, J., Holyoak, K., Nisbett, R. and Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Hornstein, N. and Lightfoot, D. (1981). *Explanation in Linguistics: The Logical Problem of Language Acquisition*. London: Longman.

Hornstein, N., Nunes, J., Grohmann, K. (eds.) (2005). *Understanding minimalism*. Cambridge: Cambridge University Press.

Howard, M. (2006) The expression of number and person through verb morphology in French interlanguage. *International Review of Applied Linguistics* 44 (1): 1-22.

Hsiao, F. and Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition* 90 (1): 3-27.

Huang, C.-T. (1982). Move wh in a language without wh movement. *Linguistic Review* 1: 369-416.

Huang, C.-T. (1995). Logical Form. In G. Webelhuth (ed.) *Government and binding theory and the minimalist program*. Oxford: Blackwell. 125-176.

Huang, C.-T., Li, A., Li, Y. (2009). *The Syntax of Chinese*. Cambridge: Cambridge University Press.

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research* 18 (3): 193–223.

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition* 27: 129-140.

Hummel, K. (2002). Second language acquisition and working memory. In F. Fabbro (ed.) *Advances in the neurolinguistics of bilingualism*. Udine: Forum. 95-118.

Hyams, N. (1986). *Language Acquisition and the Theory of Parameters*, Dordrecht: Reidel.

Hyltenstam, K. and Abrahamsson, N. (2003). Maturational constraints in SLA. In C. Doughty and M. Long (eds.) *The handbook of second language acquisition*. Oxford: Blackwell. 539-588.

Ioup, G., Boustagui, E., El Tigi, M., Moselle, M. (1994) Re-examining the critical period hypothesis: A case study of successful adult second language acquisition in a naturalistic environment. *Studies in Second Language Acquisition* 16 (1). 73-98.

Isabelli, C. (2004). Study abroad for advanced foreign language majors: optimal duration for developing complex structures. In H. Byrnes and H. Maxim (eds.) *Advanced Foreign Language Learning: A Challenge to College Programs*. Boston: Heinle and Heinle. 114-130.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research* 1146: 2-22.

James, W. (1890). *The principles of psychology*. New York: Holt.

Johnson, J. and Newport, E. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. In M. Johnson (ed.) *Brain development and cognition*. Oxford: Blackwell. 248-282.

Johnson, J. and Newport. E. (1991). Critical period effects on universal properties of language: the status of subjacency in the acquisition of a second language. *Cognition* 39: 215-258.

Johnson, J., Shenkman, K., Newport, E. and Medin, D. (1996), Indeterminacy in the grammar of adult language learners, *Journal of Memory and Language* 35: 335-352.

Jonides, J., Lewis, R., Nee, D., Lustig, C., Berman, M. and Moore, K. (2008). The mind and brain of short-term memory. *Annual Review of Psychology* 59: 193-224.

Jordens, P. (1991). Linguistic knowledge in second language acquisition. In L. Eubank (ed.) *Point counterpoint: Universal Grammar in the second language*. Amsterdam: John Benjamins. 199-218.

Juffs, A. (2001). Psycholinguistically-oriented L2 research. *Annual Review of Applied Linguistics* 21: 207-220

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society* 102 (2): 199-225.

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research* 21 (2): 121-151.

Juffs, A. and Harrington, M. (1995). Parsing effects in second language sentence processing. *Studies in Second Language Acquisition* 17 (4): 483-516.

Juffs, A. and Rodriguez, G. (2006). Second language sentence processing and working memory in college-educated and low-educated learners of English as a second language. EuroSLA, Antalya, Turkey.

Just, M. and Carpenter, P. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review* 99: 122-149.

Kail, R. and L. Hall (2001). Distinguishing short-term memory from working memory. *Memory and Cognition* 29 (1): 1-9.

Kane, M., Conway, A. and Engle, R. (1999). What do working-memory tests really measure? *Behavioral and Brain Sciences* 22 (1): 101-102.

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: evidence from metalinguistic and repair data. *Cognition* 23: 95-147.

Kim, J.-J., Kim, M., Lee, J. Lee, D. Lee, M. and Kwon, J. (2002). Dissociation of Working Memory Processing Associated with Native and Second Languages: PET Investigation. *NeuroImage* 15 (4): 879-891.

King, J. and Just, M. (1991). Individual differences in syntactic processing: The role of working memory. Journal of Memory and Language 30: 580-602.

Klein, E. and Martohardjono, G. (eds.) (1999). *Development of second language grammars: a generative approach*. Amsterdam/Philadelphia: John Benjamins.

Klein, W. and Perdue, C. (1997). The Basic Variety. *Second Language Research* 13 (4): 301-347.

Kluender, R. and Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience* 5: 196-214.

Kolers, P. (1966). Interlingual facilitation of short-term memory. *Journal of Verbal Learning and Verbal Behavior* 5: 314-319.

Köpke, B. and Nespoulous, J.-L. (2006). Working memory performance in expert and novice interpreters. *Interpreting* 8: 1-23.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Kormos, J. and Safar, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11: 261-271.

Köpke, B., Schmid, M., Keijzer, M., Dostert, S. (2007). Language attrition. Amsterdam/Philadelphia: John Benjamins.

Krashen, S. (1985). *Second language acquisition and second language learning*. Oxford, Pergamon.

Kroll, J. and de Groot, A. (eds.) (2005). *Handbook of bilingualism*. NewYork: Oxford University Press.

Lado, R. (1965). Memory span as a factor in second language learning. *International Review of Applied Linguistics* 3: 123-129.

Langacker, R. (1987). *Foundations of cognitive grammar: theoretical prerequisites*. Stanford: Stanford University Press.

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent end-state grammar. *Second Language Research* 14: 359-375.

Lardiere, D. (2007). *Ultimate attainment in second language acquisition*. New York: Routledge.

Lardiere, D. (2008) Feature Assembly in Second Language Acquisition. In J. Liceras, H. Zobl and H. Goodluck (eds.) *The role of formal features in second language acquisition.* New York, London: Lawrence Erlbaum. 107-140.

Lardiere, D. (2009) Some thoughts on a contrastive analysis of features in second language acquisition. *Second Language Research* 25 (2): 173–227.

Laurence, S. and Margolis, E. (2001). The Poverty of the stimulus argument. *British Journal for the Philosophy of Science* 52: 217-276.

Lenneberg, E. (1967). *Biological foundations of language.* New York: Wiley and Sons.

Lesaux, N. and Siegel, L. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology* 39 (6): 1005-1019.

Levelt, W. (1989). *Speaking: from intention to articulation.* Cambridge, Mass: Bradford Books.

Levelt, W. (1999). Producing spoken language: a blueprint of the speaker. In C. Brown and P. Hagoort (eds.) *The neurocognition of language.* Oxford: Oxford University Press. 83-154.

Li, X. (1998). Adult L2 Accessibility to UG: an issue revisited. In S. Flynn, G. Martohardjono and W. O'Neil (eds.) *The generative study of second language acquisition.* Mahwah, NJ: Lawrence Erlbaum. 89-110.

Lightbown, P. and Spada, N. (1993). *How languages are learned.* Oxford: Oxford University Press.

Lightfoot, D. (1999). *The development of language.* Oxford: Blackwell.

Liu, M., Schallert, D., Carroll, P. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting* 6: 19-42.

Logan, G. (2004). Working memory, task switching and executive control in the task span procedure. *Journal of Experimental Psychology, General* 133: 218-236.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie and T. Bhatia (eds.) *Handbook of second language acquisition.* San Diego: Academic Press. 413-468.

Mackey, A. (1999). Input, interaction and second language development. *Studies in Second Language Acquisition* 21: 557-587.

Mackey, A. and Gass, S. (2005). *Second language research.* Mahwah, NJ: Lawrence Erlbaum.

Mackey, A., Philp, J., Egi, T., Fujui, A., and Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (ed.) *Individual differences and instructed language learning.* Philadelphia: John Benjamins. 181-209.

MacWhinney, B. (2001). The competition model: the input, the context and the brain. In P. Robinson (ed.) *Cognition and second language instruction*. Cambridge, UK: Cambridge University Press.

MacWhinney, B. (2004). A unified model of language acquisition. In J. Kroll and A. De Groot (eds.) *Handbook of bilingualism: psycholinguistic approaches*. Oxford University Press. 49-67.

Malvern, D. Richards, B. Chipere, N. and Duran, P. (2004). *Lexical diversity and language development*. Basingstoke: Palgrave Macmillan.

Marcus, G. (1993). Negative evidence in language acquisition. *Cognition* 46: 53-85.

Marcus, G. (2001). *The algebraic mind: integrating connectionism and cognitive Science.* Cambridge, MA: MIT Press.

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research* 19 (2): 144-161.

Marinis, T., Roberts, L., Felser C. and Clahsen, H. (2005). Gaps in Second Language Sentence Processing. *Studies in Second Language Acquisition, 27,* 53-78.

Martohardjono, G. (1998). Measuring competence in L2 acquisition. In S. Flynn, G. Martohardjono and W. O'Neil (eds.) *The generative study of second language acquisition.* Mahwah, NJ: Lawrence Erlbaum.

Martohardjono, G. and Gair, J. (1993). Apparent UG inaccessibility in second language acquisition. In F. Eckman (ed.) *Confluence: linguistics, L2 acquisition and speech pathology.* Amsterdam: John Benjamins. 79-103.

Martohardjono, G. and Flynn, S. (1994). Is there an age-factor for Universal Grammar? In D. Singleton and Z. Lengyel (eds.) *The age factor in second language acquisition*. Clevedon: Multilingual Matters. 135-153.

Masoura, E. and Gathercole, S. (1999). Phonological short-term memory and foreign vocabulary learning. *International Journal of Psychology* 34: 383-388.

Masoura, E. and Gathercole, S. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory* 13(3-4): 422-429.

Matthews, G., Davies, R., Westerman, S. (2000). *Human performance: cognition, stress, and individual differences.* London: Routledge.

McBride-Chang, C. and Ho, C. (2005). Predictors of beginning reading in Chinese and English: A 2-year longitudinal study of Chinese kindergartners. Scientific Studies of Reading 9 (2): 117-144.

McClelland, J. (1989). Parallel distributed processing: Implications for cognition and development. In R. Morris (ed.) *Parallel distributed processing: implications for psychology and neurobiology*. New York: Oxford University Press. 8-45

McClelland, J. and Rumelhart, D. (1986). *Parallel distributed processing*. Cambridge, Mass: MIT Press.

McDonald, J. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language* 55 (3): 381-401.

Merrick, B. (2004). Broadening our horizons. *UK Council for International Education*. http://www.ukcisa.org.uk/files/pdf/BOHreport.pdf . Accessed 30 March 2009

Meschyan, G. and Hernandez, A. (2002). Is native-language decoding skill related to second-language learning? *Journal of Educational Psychology* 94 (1): 14-22.

Meyer, D., Glass, J., Mueller, S., Seymour, T. and Kieras D. (2001). Executive-process interactive control: a unified computational theory for answering 20 questions (and more) about cognitive ageing. *European Journal of Cognitive Psychology* 13 (1/2): 123-164.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63: 81-97.

Mitchell, R. and Myles, F. (2004). *Second language learning theories*. London: Oxford University Press.

Miyake, A., Just, M. and Carpenter, P. (1994). Working memory constraints on the resolution of lexical ambiguity. *Journal of Memory and Language* 33: 175-202.

Miyake, A. and Friedman, N. (1998). Individual differences in second language proficiency: working memory as language aptitude. In A. Healy and L. Bourne (eds.) *Foreign Language Learning*. Mahwah, NJ: Lawrence Erlbaum. 339-364.

Miyake, A. and Shah, P. (eds.) (1999). *Models of working memory*. Cambridge: Cambridge University Press.

Mizera, G. (2006). Working memory and L2 oral fluency. University of Pittsburgh (unpublished doctoral dissertation).

Mobaraki, M. (2007). Functional categories in the L2 acquisition of English morpho-syntax: a longitudinal study of two Farsi-speaking children. Durham University, UK (unpublished doctoral dissertation).

Moyer, A. (2004). *Age, accent and experience in second language acquisition*. Clevedon: Multilingual Matters.

Murphy, V. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research* 13(1): 34-65.

Myles, F. (1995). Interaction between linguistic theory and language processing in SLA. *Second Language Research* 11 (3): 235-266.

Myles, F. (2004). From data to theory: the over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society* 102 (2): 139-168.

Nani Publishing (2006). Junior High School English Student Books 1 and 2. Tainan, Taiwan: Nani Publishing.

Newmeyer, F. (1998). *Language form and language function*. Cambridge, MA: MIT Press.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science* 14 (1): 11-28.

Nicol, J. (ed.) (2001). *One mind, two languages*. Oxford: Blackwell.

Norman, D. and Shallice, T. (1986). Attention to action: willed and automatic control of behaviour. In R. Davidson, G. Schwarts and D. Shapiro (eds.) *Consciousness and self-regulation.* New York: Plenum Press. 1-18.

Norris, J. and Ortega, L. (2000). Effectiveness of L2 instruction. *Language Learning* 50 (3): 417-528.

O'Brien, I., Segalowitz, N., Collentine, J., Freed, B. (2006). Phonological memory and lexical, narrative and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics* 27: 377-402.

O'Grady, W. (2005). *Syntactic carpentry: an emergentist approach to syntax.* Mahwah, NJ: Lawrence Erlbaum.

O'Grady, W. (2008). The emergentist program. *Lingua* 118 (4): 447-464.

Ortega, L. and Iberri-Shea, G. (2005). Longitudinal research in second language acquisition. *Annual Review of Applied Linguistics* 25: 26-45.

Osaka, M. and Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society* 30 (4): 287-289.

Palladino, P. and Cornoldi, C. (2004). Working memory performance of Italian students with foreign language learning difficulties. *Learning and Individual Differences* 2004 (14): 3.

Papagno, C. and Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Journal of Experimental Psychology: Human Experimental Psychology* 48 (1): 98-107.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A. de Groot and J. Kroll (eds.) *Tutorials in bilingualism: psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum. 331-354.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam/Philadelphia: John Benjamins.

Paradis, M. (2009) *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.

Penfield, W. and Roberts, L. (1959). *Speech and brain mechanisms*. Princeton: Princeton University Press.

Perani, E., Paulesu, N., Galles, S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S., Fazio, F. and Mehler, J. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain* 121 (10): 1841-1852

Pesetksy, D. and Torrego, E. (2001). T-to-C movement: causes and consequences. In Michael Kenstowicz (ed.) *Ken Hale: a life on language*. Cambridge, MA: MIT Press.

Phillips, C. (2006). Three benchmarks for distributional models of syntax. In R. Zanuttini, H. Campos, E. Herburger and P. Portner (eds.) *Negation, tense and clausal architecture: cross-linguistic investigations*. Georgetown, DC: Georgetown University Press.

Phillips, C. and Lau, E. (2004). Foundational Issues. *Journal of Linguistics* 40: 571-591.

Pienemann, M. (1998). *Language processing and second-language development: Processability Theory*. Amsterdam: John Benjamins.

Pienemann, M. (2005). *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press

Pinker, S. (1994). *The language instinct*. New York: HarperCollins.

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: HarperCollins.

Pinker, S. (2004). Clarifying the logical problem of language acquisition. *Journal of Child Language* 31, 949–953

Pinker, S. and Prince, A. (1988) On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition* 28: 73-193.

Pinker, S. and Ullman, M. (2002) The past and future of the past tense. *Trends in Cognitive Science* 6: 456-463.

Piske, T. and Young-Scholten, M. (eds.) (2009). *Input Matters in SLA*. Clevedon: Multilingual Matters.

Platzack, C. (2001). The Vulnerable C-domain. *Brain and Language* 77, 364-377.

Pollard, C. and Sag, I. (1994). *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Poulisse, N. (1997). Language production in bilinguals. In A. de Groot and J. Kroll (eds.) *Tutorials in bilingualism: psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum. 201-224.

Prévost, P. and White, L. (2000). Accounting for morphological variation in second language acquisition: truncation or missing inflection? In M.-A. Friedemann and L. Rizzi (eds.) *The acquisition of syntax*. Harlow: Longman. 202-235.

Radford, A. (1996). Towards a structure-building model of acquisition. In H. Clahsen (ed.) *Generative perspectives on language acquisition*. Amsterdam/Philadelphia: John Benjamins. 43-90.

Radford, A. (1997). *Syntactic Theory and the Structure of English*. Cambridge, Cambridge University Press.

Radford, A. (2004). *Minimalist syntax*. Cambridge, Cambridge University Press.

Raupach, M. (1987). Production strategies in L2 performance. In H. Dechert and M. Raupach (eds.) *Psycholinguistic models of produ*ction. Norwood, NJ: Ablex.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6: 855-863.

Reis, A. and A. Castro-Caldas (1997). Illiteracy: a cause for biased cognitive development. *Journal of the International Neuropsychological Society* 3(5): 444-450.

Reis, A., Petersson, K., Castro-Caldas, A. and Ingvar, M. (2001). Formal schooling influences two- but not three-dimensional naming skills. *Brain and Cognition* 47 (3): 397-411.

Rizzi, L. (1990). *Relativized minimality*. Cambridge, Mass: MIT Press.

Rizzi, L. (1993/4) Some notes on linguistic theory and language development: the case of root infinitives. *Language Acquisition* 3: 341-393.

Rizzi, L. (1996). Residual verb second and the wh-criterion. In A. Belletti and L. Rizzi (eds.) *Parameters and functional heads*. Oxford: Oxford University Press. 63-90.

Rizzi, L. (2000) *Comparative Syntax and Language Acquisition*, Routledge, London.

Rizzi, L. (2001). On the position INT(errogative) in the Left Periphery of the Clause. In G. Cinque and G. Salvi (eds.) *Current studies in Italian syntax*. Oxford: Elsevier. 287-296.

Rizzi, L. (2005). On the grammatical basis of language development. In G. Cinque and R. Kayne (eds.) *The handbook of comparative syntax*. New York: Oxford University Press. 70-109.

Roberts, R. and Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research* 31 (6): 573-598.

Robinson, P. (1995). Attention, memory, and the noticing hypothesis. *Language Learning* 45(2): 283-331.

Robinson, P. (1996). *Consciousness, rules and instructed second language acquisition.* New York: Peter Lang.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit language learning. *Language Learning* 47: 45-99.

Robinson, P. (2001). *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Robinson, P. (ed.) (2002). *Individual differences and instructed language learning.* Amsterdam, John Benjamins.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty and M. Long (eds.) *Handbook of second language acquisition*. Oxford: Blackwell. 631-678.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics* 25: 46-76.

Rosa, E. and O'Neill, M. (1999). Explicitness, intake and the issue of awareness. *Studies in Second Language Acquisition* 21: 511-556.

Ross, J. (1967). Constraints on variables in syntax. Massachusetts Institute of Technology (doctoral thesis). Available online at http://hdl.handle.net/1721.1/15166.

Rothman, J. and Iverson, M. (2007). Input type and parameter resetting: is naturalistic input necessary? *International Review of Applied Linguistics* 45 (4): 285-320.

Rothman, J. (2008) Aspect selection in adult L2 Spanish and the Competing Systems Hypothesis: when pedagogical and linguistic rules conflict. *Languages in Contrast* 8 (1): 74-106.

Rumelhart, D. (1975). Notes on a schema for stories. In D.Bobrow and A. Collins (eds.) *Representation and understanding: studies in cognitive science*. New York: Academic Press. 185-210.

Ruschemeyer, S.-A., Fiebach, C., Kempe, V. and Friederici, A. (2005). Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping* 25: 266–286.

Sagarra, N. (2000). The longitudinal role of working memory on adult acquisition of L2 grammar. University of Illinois at Urbana-Champaign (unpublished doctoral thesis).

Sagarra, N. (2007). Working Memory and L2 processing of redundant grammatical forms. In Z. Han, E. Park, A. Revesz, C. Combs and J. Kim (eds.) *Understanding second language process.* Clevedon: Multilingual Matters. 133-147.

Saffran, J. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science* 12 (4): 110-114

Saito, S. and A. Miyake (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. *Journal of Memory and Language* 50 (4): 425-443.

Sakas, W. and Fodor, J. (2001). The structural triggers of learning. In S. Bertolo (ed.) *Language acquisition and learnability.* Cambridge: Cambridge University Press. 172–233.

Sanz, C. and Morgan-Short, K. (2005). Explicitness in pedagogical interventions: input, practice and feedback. In C. Sanz (ed.) *Mind and context in adult second language acquisition.* Washington, DC: Georgetown University Press. 234-266.

Schachter, J. (1974). An error in error analysis. *Language Learning* 24: 205-214.

Schachter, J. (1989). Testing a proposed universal. In S. Gass and J. Schachter (eds.). *Linguistic perspectives on second language acquisition.* New York: Cambridge University Press.

Schachter, J. and Yip, V. (1990). Why does anyone object to subject extraction? *Studies in Second Language Acquisition* 12: 379-392.

Schacter, D. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13: 501-518.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics* 11: 129-158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics* 13: 206-226.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: of artificial grammars and SLA. In N. Ellis (ed.) *Implicit and explicit learning of languages.* London: Academic Press. 165-210.

Schmidt, R. (2001). Attention. In P. Robinson (ed.) *Cognition and second language instruction.* Cambridge: Cambridge University Press. 3-32.

Schneider, W. and Schiffrin, R. (1977). Controlled and automatic human information processing: detection, search, and attention. *Psychological Review* 84: 1-66.

Schutze, C. (2005). Thinking about what we are asking speakers to do. In S. Kepser and M. Reis (eds.) *Linguistic evidence: Empirical, theoretical, and computational perspectives.* Berlin: Mouton de Gruyter. 457-485

Schwartz, B. (1992). Testing between UG-based and problem-solving models of L2A: developmental sequence data. *Language Acquisition* 2: 1-19.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behaviour. *Studies in Second Language Acquisition* 15: 147-163.

Schwartz, B. (1998). The second language instinct. *Lingua* 106: 133-160.

Schwartz, B. (1999). 'Transfer' and L2 acquisition of syntax: where are we now? Newcastle and Durham Working Papers in Linguistics 5: 211-234.

Schwartz, B. and Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12 (1): 40-72.

Scoville, W. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry* 20: 11–21.

Segalowitz, N. (1997). Individual differences in second language acquisition. In A. de Groot and J. Kroll (eds.) *Tutorials in bilingualism: psycholinguistic perspec*tives. Mahwah, NJ: Lawrence Erlbaum. 85-112.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (ed.) *Perspectives on fluency*. Ann Arbor, MI: University of Michigan Press. 200-219.

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty and M. Long (eds.) *The handbook of second language acquisition*. Oxford: Blackwell. 256-310.

Segalowitz, N. and Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory and Cognition* 33 (4): 644-653.

Segalowitz, N. and Segalowitz, S. (1993). Skilled performance, practice and the differentiation of speed-up from automatization effects. *Applied Psycholinguistics* 14: 369-385.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10 (3): 209-231.

Service, E. (1992). Phonology, working memory, and foreign language learning. *Quarterly Journal of Experimental Psychology, Human Experimental Psychology* 45A (1): 21-50.

Service, E. and Kohonen, V. (1995). Is the relation between phonological memory foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics* 16: 155-172.

Service, E., Simola, M., Metsanheimo, O., and Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology* 14 (3): 383-408.

Shah, P. and Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing. *Journal of Experimental Psychology: General* 125: 4-27.

Sharwood Smith, M. (1991). Speaking to many minds: on the relevance of different types of language information for the L2 learner. *Second Language Research* 7 (2): 118-132.

Sharwood Smith, M. (2008). Revisiting the role of consciousness with MOGUL. In Z. Han, E. Park, A. Revesz, C. Combs and J. Kim (eds.) U*nderstanding second language process.* Clevedon: Multilingual Matters. 1-15.

Sharwood Smith, M. and Truscott, J. (2006) Full Transfer Full Access: A Processing-Oriented Interpretation. In S. Unsworth, T. Parodi, A. Sorace and M. Young-Scholten (eds.) *Paths of Development in L1 and L2 acquisition.* Amsterdam: John Benjamins. 201-216

Skehan, P. (1989). Individual differences in second language learning. London, Arnold.

Skehan, P. (1998). *A Cognitive Approach to Language Learning.* Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (ed.) *Individual differences and instructed language learning.* Amsterdam/Philadelphia, John Benjamins. 69-95.

Skehan, P. and Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research* 1: 185-211.

Slobin, D. (ed.) (1985). *The crosslinguistic study of language acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

Smith, E. and Kosslyn, S. (2007). *Cognitive psychology: mind and brain.* New York, Prentice Hall.

Smith, N. and Tsimpli, I.-M. (1995). *The mind of a savant: language learning and modularity.* Oxford: Blackwell.

Snow, C., Arlman-Rupp, A., Hassing, Y., Jobse, J., Joosten, J., and Vorster, J. (1976). Mothers' speech in three social classes. *Journal of Psycholinguistic Research* 31: 424-444.

Sorace, A. (2003). Near-nativeness. In M. Long and C. Doughty (eds.) *Handbook of second language acquisition.* Oxford, Blackwell: 130-152.

Sorace, A. (2005). Syntactic optionality at interfaces. In L. Cornips and K. Corrigan (eds) *Syntax and variation: reconciling the biological and the social.* Amsterdam: John Benjamins. 46-111.

Sorace, A. (2006). Possible manifestations of shallow processing in advanced second language speakers. *Applied Psycholinguistics* 26: 88-91.

Sorace, A. and Keller, F. (2005). Gradience in Linguistic Data. *Lingua* 115: 1497-1524.

Speciale, G., Ellis, N. and Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics* 25 (2): 293-320.

Speidel, G. (1993). Phonological short-term memory and individual differences in learning to speak: a bilingual case study. *First Language* 13 (37): 69-91.

Spolsky, B. (1989). *Conditions for second language learning*. Oxford: Oxford University Press.

Sprouse, J., Fukuda, S. and Ono, H. (to appear). Experimental syntax, island effects, and the nature of wh-dependencies in English and Japanese. *Syntax*. Available online at: http://www.socsci.uci.edu/~jsprouse/papers/in-situ.pdf

Squire, L. (1992). Declarative and non-declarative memory: multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience* 4: 232-243.

Stromswold, K. (1995). The acquisition of subject and object wh-questions. *Language Acquisition* 4: 5-48.

Sunderman, G. and Kroll, J. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics* 30 (1): 79-100.

Swain, M. (1985). Communicative competence. In S. Gass and C. Madden (eds.) *Input in second language acquisition*. Rowley, MA: Newbury House. 235-253.

Swales, J. (1990) *Genre Analysis*. Cambridge, UK: Cambridge University Press.

Tarone, E. (1988). *Variation in interlanguage*. London, Arnold.

Tarone, E. and Bigelow, M. (2005). Impact of literacy on oral language processing: implications for second language acquisition research. *Annual Review of Applied Linguistics* 25: 77-97.

Temple, L. (1997). Memory and processing modes in language learner speech production. *Communication and Cognition* 30 (1/2): 75-90.

Thorn, A., Gathercole, S. and Frankish, C. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory. *Cognitive Psychology* 50: 133-158.

Thorn, A. and Gathercole, S. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *Quarterly Journal of Experimental Psychology A* 52 (2): 303-324.

Thornton, R. (1995). Referentiality and wh-movement in child English: juvenile *D-link*uency. *Language Acquisition* 4: 139-175.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* 74: 209-253.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition.* Cambridge, Mass, Harvard University Press.

Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn and R. Siegler (eds.) *Handbook of child psychology*. New York: Wiley.

Towell, R. (2003). Introduction: second language acquisition research in search of an interface. In R. van Hout, A. Hulk, F. Kuiken and R. Towell (eds.) *The lexicon-syntax interface in second language acquisition.* Amsterdam: John Benjamins.

Towell, R. and Dewaele, J.-M. (2005). The Role of psycholinguistic factors in the development of fluency amonst advanced learners of French. In J.-M. Dewaele (ed.) *Focus on French as a foreign language*. Clevedon: Multilingual Matters. 210-239.

Towell, R., Hawkins, R. and Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics* 17 (1): 84-119.

Trahey, M. and White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition* 16: 183-203.

Truscott, J. (2006). Optionality in Second Language Acquisition: A Generative Processing-Orientated Account. *International Review of Applied Linguistics* 44 (4): 311-330

Truscott, J and Sharwood Smith, M. (2004) Acquisition by processing: a modular perspective on language development. *Bilingualism: Language and Cognition* 7 (2): 1-20.

Truscott, J and Sharwood Smith, M. (unpublished draft manuscript) Modular growth and use of language: a processing perspective (cited with permission).

Tsimpli, I. and Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: evidence from *wh*-interrogatives in second language acquisition. *Second Language Research* 23 (2): 215-242.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Turner, M. and Engle, R. (1989). Is working memory capacity task dependent? *Journal of Memory and Language* 28: 127-154.

Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition* 4(2): 105-122.

Ullman, M. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92: 231–270.

Ullman, M. (2005). A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model. In C. Sanz (ed.) *Mind and context in adult second language acquisition*. Washington, DC: Georgetown University Press. 141-178.

Vainikka, A. (1993/4). Case in the development of English syntax. *Language Acquisition* 3: 257-325.

Vainikka, A. and Young-Scholten, M. (1994). Direct access to X' theory: Evidence from Korean and Turkish adults learning German. In T. Hoekstra and B. Schwartz (eds.) *Language acquisition studies in generative grammar*. Amsterdam: John Benjamins. 265-316.

Vainikka, A. and Young-Scholten, M. (1998). Morphosyntactic triggers in adult SLA. In M.-L. Beck (ed.) *Morphology and its interfaces in second language knowledge*. Amsterdam/Philadelphia: John Benjamins. 89-113.

Vainikka, A. and Young-Scholten, M. (2005). The roots of syntax and how they grow: Organic Grammar, the Basic Variety and Processability Theory. In S. Unsworth, T. Parodi, A. Sorace and M. Young-Scholten (eds.) *Paths of Development. in L1 and L2 Acquisition*. Amsterdam: John Benjamins. 77-106.

VanPatten, B. (1996). *Input processing and grammar instruction*. Norwood, NJ: Ablex.

VanPatten, B. (2000). Thirty years of input (or intake, the neglected sibling). In B. Swierzbin, F. Morris, M. Anderson, C. Klee and E. Tarone (eds.) *Social and cognitive factors in second language acquisition*. Somerville, MA: Cascadilla Press.287-311.

VanPatten, B. (2004). *Processing instruction: theory, research and commentary*. Mahwah, NJ: Lawrence Erlbaum.

VanPatten, B. (2005). Processing instruction. In C. Sanz (ed.) *Mind and context in adult second language acquisition*. Washington, DC: Georgetown University Press. 267-282.

Van Valin, R. (1998). The acquisition of wh-questions and the mechanisms of language acquisition. In M. Tomasello (ed.) *The new psychology of language*. 221-251.

Vos, S., Gunter, T., Schriefers, H., Friederici, A. (2001). Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes* 16 (1): 65-103.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics* 25 (3): 315-339.

Walters, J. (2005). *Bilingualism*. Mahwah, NJ: Lawrence Erlbaum.

Watanabe, Y. and Bergsleithner, J. (2007). Working Memory Measurements in SLA: A Research Synthesis and Meta-analysis. Presentation at American Association of Applied Linguistics, Costa Mesa, California.

Waters, G. and Caplan, D. (1996). The capacity theory of sentence comprehension *Psychological Review* 103(4): 761-772.

Waters, G. and Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments and Computers* 35 (4): 550-564.

Waters, G. and Caplan, D. (2004). Verbal working memory and online syntactic processing: evidence from self-paced listening. *Quarterly Journal of Experimental Psychology A* 57 (1): 129-163.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—3rd Edition (WAIS-3®)*. San Antonio: Harcourt Assessment.

Wexler, K. (1994). Finiteness and head movement in early child grammar. In D. Lightfoot and N. Hornstein (eds.), *Verb movement*. Cambridge: Cambridge University Press. 305-350.

White, L. (1989). *Universal Grammar and second language acquisition*. Amsterdam, John Benjamins.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge, Cambridge University Press.

White, L. and F. Genesee (1996). How native is near native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research* 11: 233-265.

White, L. and A. Juffs (1998). Constraints on wh-movement in two different contexts of nonnative language acquisition: competence and processing. In S. Flynn, G. Martohardjono and W. O'Neil (eds.) *The generative study of second language acquisition.* Mahwah, NJ: Lawrence Erlbaum. 111-129.

Whong-Barr, M. and Schwartz, B. (2002). Morphological and syntactic transfer in child L2 acquisition of the dative alternation. Studies in Second Language Acquisition 24: 579-616.

Williams, J. (1999). Memory, attention and inductive learning. *Studies in Second Language Acquisition* 21: 1-48.

Williams, J. (2005). Associationism and connectionism. In K.Brown (ed.) *Encyclopedia of language and linguistics*. Elsevier: Oxford.

Williams, J. and Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning* 53 (1): 67-121.

Winke, P. (2005) Individual differences in adult Chinese second language acquisition: The relationships among aptitude, memory and strategies for learning. Georgetown University, Washington, DC (unpublished doctoral thesis).

Wright, C. (2006). Is interlanguage variation in complex CP structures predictable? *Proceedings of the Fourth University of Cambridge Postgraduate Conference in Language Research.* Cambridge, UK: Cambridge Institute of Language Research. 277-284.

Yip, V. (1995). *Interlanguage and learnability: from Chinese to English.* Amsterdam: John Benjamins.

Yoshimura, Y. (2001). The role of working memory in language aptitude. In X. Bonch-Bruevich, W. Crawford, J. Hellerman, C. Higgins and H. Nguyen (eds.) *The past, present and future of second language research: Proceedings of Second Language Research Forum.* Somerville, MA, Cascadilla Press: 144-163.

Young-Scholten, M., Ijuin, C. and Vainikka, A. (2005). Organic Grammar as a measure of development. TESOL meeting. San Antonio, Texas.

Yuan, F. and Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics* 24(1): 1-27.

APPENDICES

Appendix A: Materials used in the first study

A i: Questionnaire and permission form

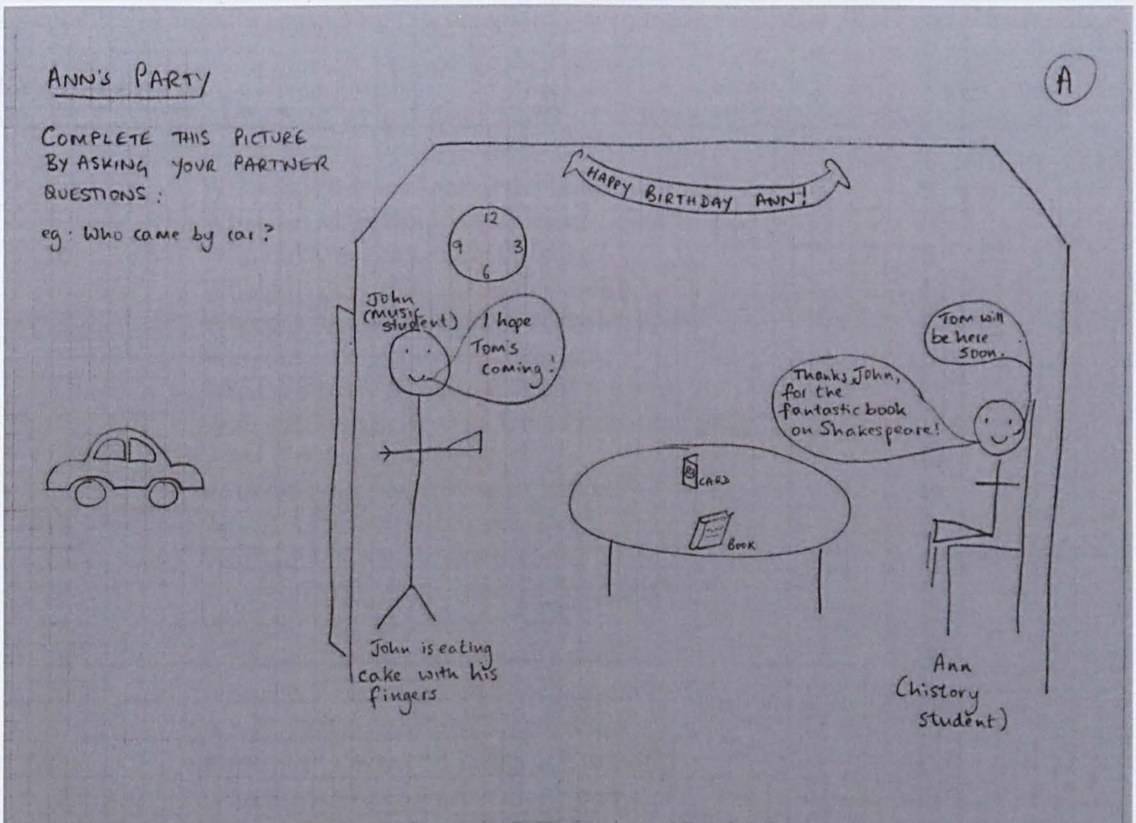
What's your name?		
Where are you from?		
How old were you when you started learning English?		
How old are you now? How many years have you been using English?		
How much English did you hear outside school, if any (eg films, pop music, English clubs)?	1 = <2 hours per month 2 = 2- 8 hours per month 3 = >2 hours per week	*Please write 1, 2 or 3 below*
How long have you been in the UK?	1 = < 1 month 2 = 1-3 months 3 = > 3 months	*Please write 1, 2 or 3 below*
How much English are you using at the moment?	1 = <5 hours a day 2 = 5-10 hours a day 3 = > 10 hours a day	*Please write 1, 2 or 3 below*
What was your latest IELTS score or equivalent, and what year was it?	Total: Reading Writing	Year: Listening Speaking
Do you speak any other languages? If so, please list them.		
How many years have you been using them?	1 = <1 year 2 = 1-3 years 3 = >more than 3 years	*Please write 1, 2 or 3 below*

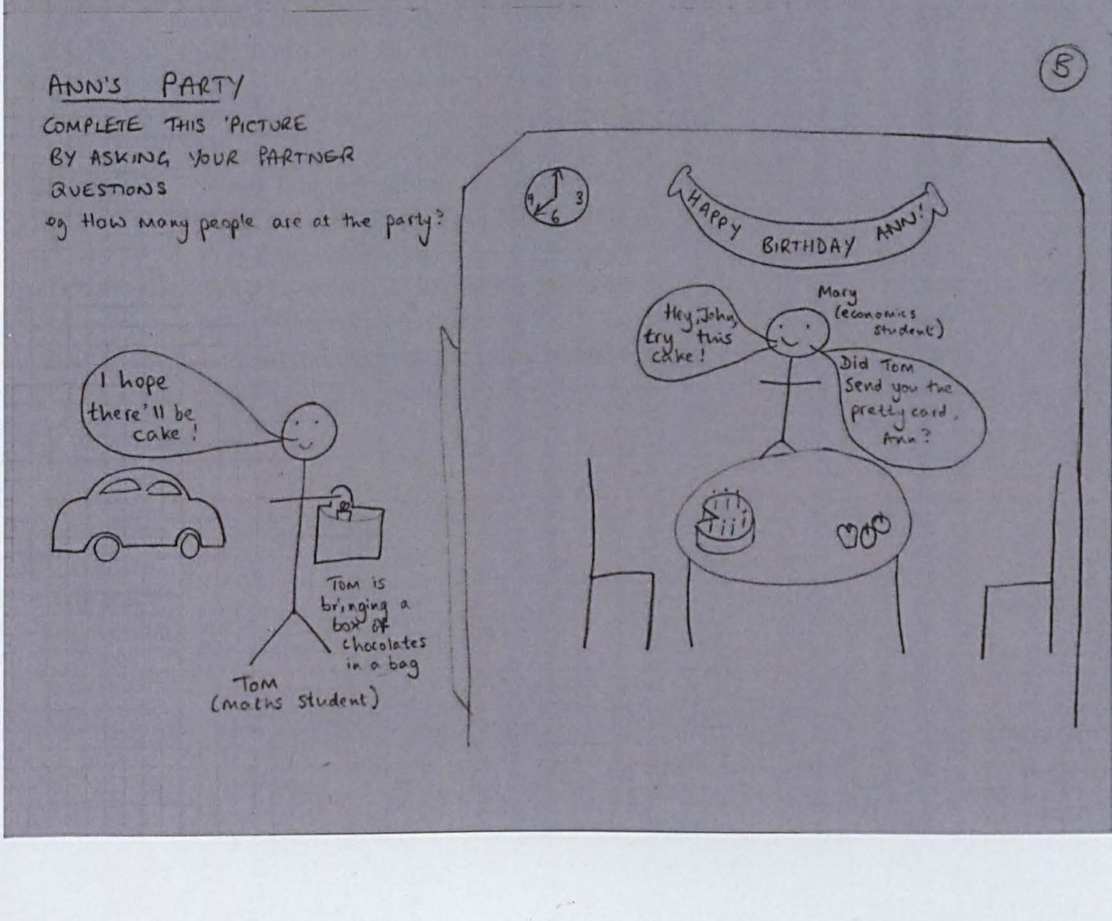
Thank you for taking part in this research project for Newcastle University. This information is confidential and your name will not be used. Please sign below to show that you are happy for your data to be published for research purposes.

Signature: *Date:*

Linguistic Tasks

A ii: Oral question elicitation task – version A and B used at Time 1





A iii: Grammaticality judgement task (mark up version, grouped by type)

Look at the following questions and mark how grammatically acceptable you think they are: completely unacceptable (-3), usually unacceptable (-2), possibly unacceptable (-1), possibly acceptable (1), usually acceptable (2), completely acceptable (3). Mark (0) if you are not sure.

EXAMPLE:
What did Ann think the book of was from John? -3 What did Mary see John was eating? 3

3	What did Ann say looked delicious?	S
7	What did Mary think tasted nice?	S
10	Who did Mary say made the cake?	S
16	Who did Mary think liked the book?	S
13	Who did Ann expect to arrive soon?	S nf
19	Who did John believe sent the card?	S nf
5	What did Mary think John ate at the party?	O
17	Who did Tom believe he would see at the party?	O
23	What did Ann say she liked?	O
25	Who did Ann say she expected soon?	O
18	Who did Ann expect to see later?	O nf
22	Who did Mary hope to see soon?	O nf
20	What did Mary think it tasted delicious?	S*
21	Who did John hope arrive late?	S*
28	Who did John like come to the party?	S*
30	Who did Tom expect give the present?	S*
1	What did Mary see was John eating?	O*
9	What did Tom say he bring as a present?	O*
12	Who did Ann expect saw at the party?	O*
15	What did John know did Ann like?	O*
2	What did John eat after he arrived?	Adj
6	Who did Ann thank before she expected Tom?	Adj
8	What did Tom bring a present after he sent?	Adj*
14	What did Mary see the card while John ate?	Adj*
11	What Tom brought to the party?	D
27	Why did Ann think Tom arrived late at the party?	D
29	Did John believe the cake was tasty?	D
4	Why Tom arrived late at the party?	D*
24	Why Tom come to the party?	D*
26	Did John think was the book interesting?	D*

Key:
S = Subject
O = Object
nf = non-finite
Adj = Adjunct
D = distracter
* = ungrammatical

A iv: Question Formation task (mark-up version, sorted by type)

Look at the following sentences. Please form a grammatical question beginning with a wh-word ("why, who, how, when, where, what") that questions the underlined phrase.

Example:
Mary said that she liked the card.
Possible answer - What did Mary say she liked?
Not possible – Mary said that she liked what?

4	John knew that Ann liked books about Shakespeare.	"That" trace (*)
2	Ann thanked John before she expected Tom to arrive.	Adj (*)
9	Tom arrived at the party after he parked his car.	Adj (*)
15	Tom sent a card before he brought a present.	Adj (*)
3	Books about Shakespeare made Ann happy.	Comp NP (*)
14	The student of maths brought a present in a bag.	Comp NP (*)
11	The student of music ate cake with his fingers.	PP (*)
5	Ann said that she was expecting Tom.	O
10	Mary said she liked the card.	O
13	Mary thought John enjoyed the cake.	O
1	Ann knew Tom had sent her a card.	S
6	Ann thought Tom would come later.	S
7	Mary saw Ann had put the apples on the table.	S
8	John expected Tom to come soon.	S
12	John thought he would enjoy the cake.	S

Key:
S = Subject
O = Object
Adj = Adjunct
* = subjacency constraint on extraction

WM tasks

A v: Digits Back

(L2 English shown) – used in both first and second study
(same instructions translated into Mandarin and used in the Mandarin digital sound file, with different digits)

Instructions: listen to these sets of numbers, I am going to say some numbers, and when I've finished, I want you to repeat them back to me in reverse order, so if I say 1, 2, 3, you say, 3, 2, 1.We'll start with 3, and if that's no problem we'll then try 4, then 5 until it gets too hard. (Numbers to be read clearly at 1 digit per second.)

Now let's practice: I'll say:
2 3 4 - you say?
1 4 2

Set 1

1	5	9
3	2	7
4	7	1

Set 2

4	7	2	3
1	6	4	8
1	8	6	3

Set 3

1	3	5	2	4
9	5	7	3	2
1	2	4	5	8

Set 3

1	4	7	5	8	6
2	6	4	8	6	1
3	4	6	7	9	8

Set 4

2	6	9	8	4	5	7
6	3	9	6	8	5	1
4	6	1	3	7	9	5

A vi: Story Recall Text and Marking Scheme
(for English L2) used in both first and second study

1. Sentence recall practice examples:
Instruction
"Let's practice recalling some short texts – these are all short stories about some sad things that can happen to students" (change final word to fit context if necessary, eg "friends of mine").

Someone stole my car yesterday, so I went to the police.
Angela has lost all her money. She doesn't know if she should go to the police or phone her parents.
Someone took Mary's computer and all her books last night, so she's feeling terrible. She tried to stop the thief but he escaped.

2. Story Recall text:
Instruction
"Now I'm going to tell you a story about my friend Andrew Wright. (*Repeat*) Andrew Wright. This is a happy story this time. When I've finished, I want you to tell it back to me exactly as you heard it, as far as possible using the same words and phrases". *To be read naturally (or use audio file for multiple tests to maintain consistency of presentation); should take around 32 seconds.*

Andrew Wright
was sitting in the park
talking to his girlfriend
when he noticed a small bag
lying on the ground.
He picked it up to see what was inside it
and discovered a twenty-pound note
and some coins.
Although he tried to see
if someone was looking for it,
he could not find anyone in the park.
When he asked his girlfriend
what he should do,
she told him to keep it
and buy her some new clothes.

Marking scheme: total score /50 – used as main score
Meaning = 1 per "schematic episode" as shown in merged cell.
Accuracy = 1 per semantic and syntactically accurate target unit as underlined; some leeway or half-points for near-match of lexis or syntax.

		Meaning (/10)	Accuracy (/40)
	Andrew Wright		1
1	was sitting in the park	1	2
2	talking to his girlfriend	1	3
3	when he noticed a small bag	1	3
	lying on the ground.		3
4	He picked it up to see what was inside it	1	3
	and discovered a twenty-pound note		3
5	and some coins.	1	1
	Although he tried to find out		3
6	if the owner was looking for it,	1	3
7	he could not see anyone in the park.	1	3
	When he asked his girlfriend		3
8	what he should do,	1	3
9	she told him to keep it	1	3
10	and buy her some chocolate and flowers for her.	1	3
	Total (/50)	(/10)	(/40)

Example marked up (genuine data):

	Meaning	Accuracy
When Andrew er (surname omitted)		.5
Sitting er was sitting in the park	1	2
with her girlfriend (no talk vb, no progressive, wrong gender on pronoun; half-point for lexical item *girlfriend*, 0 if *friend*)	1	.5
He saw a bag (*bag* = 1)		
lay on the chair (lexis for *lie*+ with tense marked = 1)	1	2
and he pick it up (lexis but no subordination, no tense)	1	1
and s/ find there was twenty pound (lexis)		1
and some coins in it. (lexis, prepositional phrase)	1	2
She He want to find (use of non-finite)		1
the owners (lexis)	1	1
but there was no <> there was no in the park. (existential subject, prepositional phrase)	1	2
Sh[e] He don't know he doesn't er sh[e] he he doesn't know what she could do (embedded qu, but no tense on matrix verb, gender error on pronoun)	1	1
And her girlfriend er her girlfriend ask him to keep it (correct phrase structure, no tense on matrix)	1	2
And br[ought] and brought her some new clothes. (idea of some kind of offer – *brought* probably phonetic mismatch for *bought*)	1	1
Total = 27 (10 + 17)	10	17

Adapted from the Adult Memory and Information Processing Battery (AMIPB) test manual (Coughlan and Hollows 1985).

A vii: Listening Span
(incorporating Word Span and Sentence Span) used in first study

Practice:
*Turn to the **left.***

*Turn **right** ahead.* Recall: left, right

Turn **left** at the end. (5 words)

Go straight **down** the street.

At the metro station turn **right**.

Go **right** after the traffic lights.

Walk **up** the street until the lights.

Take the second turn on the **left**.

Go straight **down** the road to the corner.

Take the first **right** after the car park.

Turn **right** after the bridge but before the metro.

Walk **down** this hill until you pass the pub.

Go all the way past the bookshop and turn **left**.

Walk **down** this path till you see the bus station.

After the library walk all the way **down** to the end.

Go **up** the hill until you see the bus stop ahead.

Walk straight along this road then turn **right** at the bus station. (12 words)

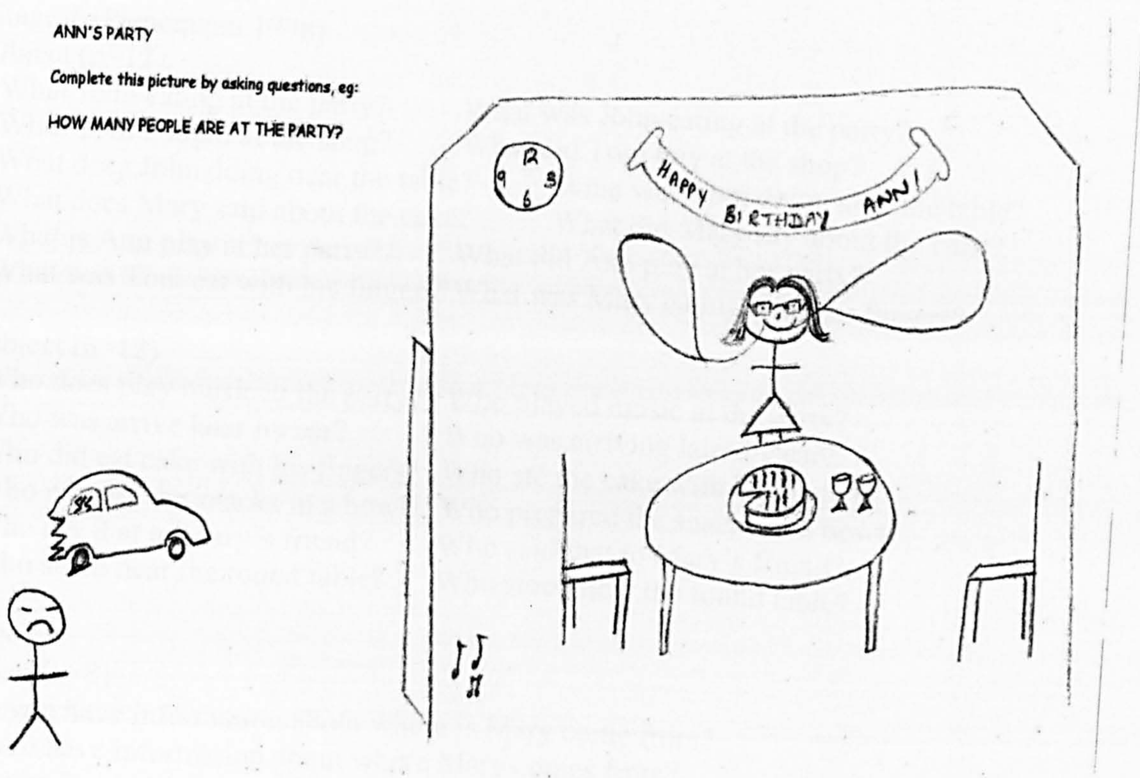
Go all the way past the shops and take the second **left**.

Appendix B: Material revised for use in second study

B i. Revised questionnaire and permission form

What's your name?		
Where are you from?		
How old were you when you started learning English?		
How old are you now? How many years have you been using English?		
How much English did you hear outside school, if any (eg films, pop music, English clubs)?	1 = <2 hours per month 2 = 2- 8 hours per month 3 = >2 hours per week	Please write 1, 2 or 3 below
How long have you been in the UK?	1 = < 1 month 2 = 1-3 months 3 = > 3 months	Please write 1, 2 or 3 below
How much English are you using at the moment?	1 = <5 hours a day 2 = 5-10 hours a day 3 = > 10 hours a day	Please write 1, 2 or 3 below
What was your latest IELTS score or equivalent, and what year was it?	Total: Reading Writing	Year: Listening Speaking
Have you had a break in using English since school? If so, how long?	Yes/No – delete as appropriate. If Yes: Number of years -	
Do you speak any other languages? If so, please list them.		
How many years have you been using them?	1 = less than 1 year 2 = 1-3 years 3 = more than 3 years	Please write 1, 2 or 3 below
Thank you for taking part in this research project for Newcastle University. This information is confidential and your name will not be used. Please sign below to show that you are happy for your data to be published for research purposes. *Signature:* *Date:*		

B ii. Oral elicitation task (participant's picture)



B iii: Timed Grammaticality Judgement data (DMDX test stimuli)

Short-distance movement (adapted from first study)

Stage 5 (Pienemann 1998)
Object (n=12)
*What John eating at the party? What was John eating at the party?
*What Tom bought at the shop? What did Tom buy at the shop?
*What does John doing near the table? What was Tom doing near the table?
*What does Mary said about the cake? What did Mary say about the cake?
*What is Ann play at her party? What did Ann play at her party?
*What was Tom eat with his fingers? What was Mary eating with her fingers?

Subject (n=12)
*Who does play music at the party? Who played music at the party?
*Who was arrive later by car? Who was arriving later by car?
*Who did eat cake with his fingers? Who ate the cake with his fingers?
*Who prepare the snacks in a bowl? Who prepared the snacks in a bowl?
*Who say that to Mary's friend? Who said that to Mary's friend?
*Who stand near the round table? Who stood near the round table?

Stage 6
Object (n=8)
*Do you have information about where is Mary come from?
Do you have information about where Mary comes from?
*Do you know this party is going to celebrate some who?
Do you know who this party is going to celebrate?
*Did Ann know what did John eat? Did Mary know what John ate?
*Do you know who was Ann expecting Do you know who John was expecting?

Subject/copula complement (n=8)
*Do you see was who eating cake? Do you know who was eating cake?
*Do you know who coming by car? Do you know who is coming by car?
*Did Mary know who does like Shakespeare? Did Ann know who likes cake?
*Did John think was the cake tasty? Did Mary think the card was pretty?

Long-distance movement (based on Juffs and Harrington 1995; White and Juffs 1998)
Ungrammatical subjacency violation (12)

Subject islands:
*What did books about make Ann happy?
*Who did a story by please Ann?
*What did the news about surprise John?
*What was a dish of cooked by Mary?

Relative clause islands:
*What did Mary see the man who stole?
*Who did Tom meet the girl who married?
*What did Ann make the cake which contained?
*Who did John love the girl who married?

Adjunct islands:
*Who did Ann thank John after she saw?
*What did Tom bring a present after he sent?
*What did Mary see the card while John ate?
*What did Ann get excited because Tom brought?

Grammatical (16)
Subject extraction (finite)
Who did Ann say liked her friend?
What did John think crashed into the car?
Who do you suppose wanted to marry Ann?
Who did Mary announce would be the new teacher?

Object extraction (finite)
Who did Mary say her friend liked?
What did John think the car crashed into?
Who do you suppose John wanted to marry?
Who did Mary announce the new teacher would be?

Subject extraction (infinitival)
Who did Ann want to win the game?
Who did Tom expect to beat Mary?
Who did Ann expect to go with John?
Who did John believe to be the best player?

Object extraction (infinitival)
What did John want to win?
Who did Tom expect to beat?
Who did Mary expect to go with?
Who did John believe the best player to be?

distracter tokens:
Did the student who worked hard do well?
Did the boy who practised the piano pass his exam?
*What did Mary think it tasted delicious?
*Who did Ann see she brought flowers?
*What did John know Ann liked it?
*Who did Tom expect he would see her?
*Did the student worked hard do well?
*Did the girl practised the piano pass her exam?
*Who did Mary expect saw at the party?
*What did Tom say he bring as a present?
*What did John know did Ann like?
*Who did John like come to the party?
*Who did John hope arrive later?
*Who did Tom expect give the present?
*Who did Ann know send her a card?
*What is the book about on the table?

B iv: Listening Span task - Revised directions task used in second study

Instructions*: Listen to these directions. We will start with two directions. You will have to remember the final place at the end of each direction and tell me these words at the end of the group of directions.*

Practice for final word recall, listen to both sentences first:
Turn left by the lights.
Walk right down the street **Final words?**

But you also have to listen carefully to the direction and tell me straightaway after each sentence what direction word you can hear – left, right, up or down.

Practice for direction word plus final word recall.
Walk up the hill to the bookshop. ***Direction word***?
Turn left before the library.

Now, what were the 2 final words or places?

Test stimuli
Sets of 2
Go down the road to the corner.
Walk up the street until the cafe.

Turn right after the book shop.
Turn left after the train station.

Go down the hill to the park.
Take the next left turn by the shop.

Sets of 3
Keep going up after the lights.
Turn left after the post office.
Go down until the museum.

Take the second right by the park.
Take the first right after the car park.
Walk straight down until the traffic lights.

Walk down this hill until the shop.
Turn left before the swimming pool.
Go up the road until the bridge.

Sets of 4
Take the second left by the café.
Turn right after the supermarket.
Go all the way down past the shops.
Walk up the street until the metro.

Take the first left by the library.
Go straight down to the bus station.
Keep going up this hill to the park.
Keep going down until the library.

Go left past the supermarket.
Turn down just after the cafe.
Go up the hill at the traffic lights.
Walk straight down past the car park.

Sets of 5
Take the first left after the book shop.
Walk straight up until the train station.
Walk down this road until the metro.
Turn right before the bakery.
Walk straight down along the main road.

Go up this path until the metro.
Walk down the street until the lights.
Turn right after the bus stop.
Take the first right after the cafe.
Go down here until the supermarket.

Take the first right by the car park.
Go straight down to the bus station.
Keep going up this hill to the bank.
Take the first right after the shop.
Walk straight up until the traffic lights.

Appendix C: Tables giving full linguistic and correlational data for first study

Table C i: Grammaticality judgement data, Time 1 to Time 3 (Mean, SD and Range)

	Time 1	Time 2	Time 3	Change by Time 3
Mean	8.36	7.82	10.45	2.09
Std. Deviation	3.32	3.89	3.26	1.92
Minimum	4	3	5	-1
Maximum	16	16	16	5

Table C ii: Question formation data, Time 1 to Time 3 (Mean, SD and Range)

	Time 1	Time 2	Time 3	Change by Time 3
Mean	7.55	10.09	10.36	2.82
Std. Deviation	3.96	2.74	2.94	4.07
Minimum	1	4	6	-4
Maximum	13	13	15	9

Correlational tables C iii and C iv follow.

Table C iii: Correlations between oral measures (question total and question ratio) and WM measures at Time 1, Time 3 and with change by Time 3

		question total time 1	question total time 3	change question total	question ratio	question ratio time 3	change question ratio	DBL1	DBL2	SRL1	SRL2	Word span	Sentence span
Variable co-reference		1	2	3	4	5	6	7	8	9	10	11	12
1	Correlation Coefficient	1.000	-.516	-.530	.805(**)	-.519	-.630(*)	-.199	.187	.261	.311	.276	-.014
	Sig. (2-tailed)	.	.104	.094	.003	.102	.038	.558	.605	.467	.352	.411	.968
	N	11	11	11	11	11	11	11	10	10	11	11	11
2	Correlation Coefficient	-.516	1.000	.995(**)	-.449	.765(**)	.865(**)	-.117	-.169	.034	-.346	-.002	.221
	Sig. (2-tailed)	.104	.	.000	.166	.006	.001	.731	.641	.926	.297	.995	.514
	N	11	11	11	11	11	11	11	10	10	11	11	11
3	Correlation Coefficient	-.530	.995(**)	1.000	-.422	.794(**)	.856(**)	-.066	-.146	.079	-.299	.007	.224
	Sig. (2-tailed)	.094	.000	.	.196	.004	.001	.846	.687	.828	.372	.984	.507
	N	11	11	11	11	11	11	11	10	10	11	11	11
4	Correlation Coefficient	.805(**)	-.449	-.422	1.000	-.190	-.556	.159	.564	.589	.543	.411	.081
	Sig. (2-tailed)	.003	.166	.196	.	.576	.075	.640	.089	.073	.084	.210	.813
	N	11	11	11	11	11	11	11	10	10	11	11	11
5	Correlation Coefficient	-.519	.765(**)	.794(**)	-.190	1.000	.833(**)	.253	.289	.403	.065	.363	.561
	Sig. (2-tailed)	.102	.006	.004	.576	.	.001	.453	.418	.248	.850	.273	.073
	N	11	11	11	11	11	11	11	10	10	11	11	11
6	Correlation Coefficient	-.630(*)	.865(**)	.856(**)	-.556	.833(**)	1.000	-.046	-.024	-.171	-.422	.183	.434
	Sig. (2-tailed)	.038	.001	.001	.075	.001	.	.894	.947	.637	.196	.589	.183
	N	11	11	11	11	11	11	11	10	10	11	11	11
7	Correlation Coefficient	-.199	-.117	-.066	.159	.253	-.046	1.000	.325	.442	.355	.447	.450
	Sig. (2-tailed)	.558	.731	.846	.640	.453	.894	.	.359	.201	.284	.168	.165
	N	11	11	11	11	11	11	11	10	10	11	11	11
8	Correlation Coefficient	.187	-.169	-.146	.564	.289	-.024	.325	1.000	.304	.352	.557	.212
	Sig. (2-tailed)	.605	.641	.687	.089	.418	.947	.359	.	.394	.319	.094	.557
	N	10	10	10	10	10	10	10	10	10	10	10	10
9	Correlation Coefficient	.261	.034	.079	.589	.403	-.171	.442	.304	1.000	.802(**)	.329	.393
	Sig. (2-tailed)	.467	.926	.828	.073	.248	.637	.201	.394	.	.005	.353	.262
	N	10	10	10	10	10	10	10	10	10	10	10	10
10	Correlation Coefficient	.311	-.346	-.299	.543	.065	-.422	.355	.352	.802(**)	1.000	.218	.074
	Sig. (2-tailed)	.352	.297	.372	.084	.850	.196	.284	.319	.005	.	.520	.829
	N	11	11	11	11	11	11	11	10	10	11	11	11
11	Correlation Coefficient	.276	-.002	.007	.411	.363	.183	.447	.557	.329	.218	1.000	.767(**)
	Sig. (2-tailed)	.411	.995	.984	.210	.273	.589	.168	.094	.353	.520	.	.006
	N	11	11	11	11	11	11	11	10	10	11	11	11
12	Correlation Coefficient	-.014	.221	.224	.081	.561	.434	.450	.212	.393	.074	.767(**)	1.000
	Sig. (2-tailed)	.968	.514	.507	.813	.073	.183	.165	.557	.262	.829	.006	.
	N	11	11	11	11	11	11	11	10	10	11	11	11

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table C iv: Correlations between grammaticality judgement task scores (by submeasure at Time 1 and Time 3) with WM measures

		grammatical time 1	ungrammatical time 1	grammatical time 3	ungrammatical time 3	objects time 1	subjects time 1	objects time 3	subjects time 3	DB L1	DB L2	SR L1	SR L2	Word span	Sentence span
Variable coreference		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Correlation Coefficient	1.000	-.093	-.124	.684(*)	.858(**)	.466	-.146	.419	-.012	.310	-.186	-.125	.429	.513
	Sig. (2-tailed)	.	.786	.715	.020	.001	.148	.668	.199	.973	.384	.607	.714	.188	.107
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
2	Correlation Coefficient	-.093	1.000	.461	.469	.178	.316	.368	.570	.199	.083	.516	.375	-.083	.007
	Sig. (2-tailed)	.786	.	.154	.146	.600	.344	.266	.067	.558	.820	.127	.256	.809	.983
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
3	Correlation Coefficient	-.124	.461	1.000	.239	.232	-.407	.932(**)	.411	.198	-.068	-.003	.262	.016	-.107
	Sig. (2-tailed)	.715	.154	.	.478	.492	.214	.000	.209	.559	.851	.993	.436	.962	.754
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
4	Correlation Coefficient	.684(*)	.469	.239	1.000	.809(**)	.499	.195	.791(**)	.351	.511	.220	.042	.476	.540
	Sig. (2-tailed)	.020	.146	.478	.	.003	.118	.566	.004	.290	.131	.542	.903	.139	.086
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
5	Correlation Coefficient	.858(**)	.178	.232	.809(**)	1.000	.199	.143	.660(*)	.207	.458	.049	.179	.382	.339
	Sig. (2-tailed)	.001	.600	.492	.003	.	.557	.676	.027	.541	.183	.893	.599	.246	.307
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
6	Correlation Coefficient	.466	.316	-.407	.499	.199	1.000	-.409	.329	-.137	.385	.050	-.223	.267	.444
	Sig. (2-tailed)	.148	.344	.214	.118	.557	.	.211	.323	.689	.272	.892	.509	.427	.171
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
7	Correlation Coefficient	-.146	.368	.932(**)	.195	.143	-.409	1.000	.285	.214	-.169	-.009	.157	-.021	-.053
	Sig. (2-tailed)	.668	.266	.000	.566	.676	.211	.	.395	.527	.640	.980	.644	.952	.877
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
8	Correlation Coefficient	.419	.570	.411	.791(**)	.660(*)	.329	.285	1.000	.201	.470	-.029	.048	.349	.080
	Sig. (2-tailed)	.199	.067	.209	.004	.027	.323	.395	.	.553	.170	.937	.890	.292	.814
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11

Correlations, cont'd		grammatical time 1	ungrammatical time 1	grammatical time 3	ungrammatical time 3	objects time 1	subjects time 1	objects time 3	subjects time 3	DB L1	DB L2	SR L1	SR L2	Word span	Sentence span
Variable coreference		1	2	3	4	5	6	7	8	9	10	11	12	13	14
9	Correlation Coefficient	-.012	.199	.198	.351	.207	-.137	.214	.201	1.000	.325	.442	.355	.447	.450
	Sig. (2-tailed)	.973	.558	.559	.290	.541	.689	.527	.553	.	.359	.201	.284	.168	.165
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
10	Correlation Coefficient	.310	.083	-.068	.511	.458	.385	-.169	.470	.325	1.000	.304	.352	.557	.212
	Sig. (2-tailed)	.384	.820	.851	.131	.183	.272	.640	.170	.359	.	.394	.319	.094	.557
	N	10	10	10	10	10	10	10	10	10	10	10	10	10	10
11	Correlation Coefficient	-.186	.516	-.003	.220	.049	.050	-.009	-.029	.442	.304	1.000	.802(**)	.329	.393
	Sig. (2-tailed)	.607	.127	.993	.542	.893	.892	.980	.937	.201	.394	.	.005	.353	.262
	N	10	10	10	10	10	10	10	10	10	10	10	10	10	10
12	Correlation Coefficient	-.125	.375	.262	.042	.179	-.223	.157	.048	.355	.352	.802(**)	1.000	.218	.074
	Sig. (2-tailed)	.714	.256	.436	.903	.599	.509	.644	.890	.284	.319	.005	.	.520	.829
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
13	Correlation Coefficient	.429	-.083	.016	.476	.382	.267	-.021	.349	.447	.557	.329	.218	1.000	.767(**)
	Sig. (2-tailed)	.188	.809	.962	.139	.246	.427	.952	.292	.168	.094	.353	.520	.	.006
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11
14	Correlation Coefficient	.513	.007	-.107	.540	.339	.444	-.053	.080	.450	.212	.393	.074	.767(**)	1.000
	Sig. (2-tailed)	.107	.983	.754	.086	.307	.171	.877	.814	.165	.557	.262	.829	.006	.
	N	11	11	11	11	11	11	11	11	11	10	10	11	11	11

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Appendix D: Tables giving full linguistic and correlational data for second study

Table D i: Non-parametric correlations (Spearman's rho) between WM tests at Time 1 and at Time 2
(a) for WM at Time 1

Time 1		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2	Listening Span Time 2	Digits Back L1 Time 2	Digits Back L2 Time 2	Story Recall L1 Time 2	Story Recall L2 Time 2
Listening Span	Correlation Coefficient	1.000	.131	.154	-.177	-.300	.185	.171	.159	.044	.190
	Sig. (2-tailed)	.	.489	.426	.367	.101	.320	.358	.393	.813	.306
	N	31	30	29	28	31	31	31	31	31	31
Digits Back L1	Correlation Coefficient	.131	1.000	.569(**)	-.030	-.109	.173	.668(**)	.021	-.034	-.110
	Sig. (2-tailed)	.489	.	.001	.876	.559	.352	.000	.909	.858	.556
	N	30	31	29	29	31	31	31	31	31	31
Digits Back L2	Correlation Coefficient	.154	.569(**)	1.000	.043	.050	-.210	.343	.585(**)	.372(*)	.126
	Sig. (2-tailed)	.426	.001	.	.830	.792	.265	.063	.001	.043	.509
	N	29	29	30	27	30	30	30	30	30	30
Story Recall L1	Correlation Coefficient	-.177	-.030	.043	1.000	.402(*)	-.075	-.112	-.030	.378(*)	.321
	Sig. (2-tailed)	.367	.876	.830	.	.031	.699	.563	.876	.043	.089
	N	28	29	27	29	29	29	29	29	29	29
Story Recall L2	Correlation Coefficient	-.300	-.109	.050	.402(*)	1.000	.054	-.135	-.154	.253	.474(**)
	Sig. (2-tailed)	.101	.559	.792	.031	.	.768	.462	.401	.162	.006
	N	31	31	30	29	32	32	32	32	32	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D i: Non-parametric correlations (Spearman's rho) between WM tests at Time 1 and at Time 2 (b) for WM at Time 2

Time 2		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2	Listening Span Time 2	Digits Back L1 Time 2	Digits Back L2 Time 2	Story Recall L1 Time 2	Story Recall L2 Time 2
Listening Span Time 2	Correlation Coefficient	.185	.173	-.210	-.075	.054	1.000	.347	.090	-.090	.080
	Sig. (2-tailed)	.320	.352	.265	.699	.768	.	.051	.626	.624	.664
	N	31	31	30	29	32	32	32	32	32	32
Digits Back L1 Time 2	Correlation Coefficient	.171	.668(**)	.343	-.112	-.135	.347	1.000	.171	-.098	.049
	Sig. (2-tailed)	.358	.000	.063	.563	.462	.051	.	.350	.593	.790
	N	31	31	30	29	32	32	32	32	32	32
Digits Back L2 Time 2	Correlation Coefficient	.159	.021	.585(**)	-.030	-.154	.090	.171	1.000	.475(**)	.188
	Sig. (2-tailed)	.393	.909	.001	.876	.401	.626	.350	.	.006	.303
	N	31	31	30	29	32	32	32	32	32	32
Story Recall L1 Time 2	Correlation Coefficient	.044	-.034	.372(*)	.378(*)	.253	-.090	-.098	.475(**)	1.000	.054
	Sig. (2-tailed)	.813	.858	.043	.043	.162	.624	.593	.006	.	.771
	N	31	31	30	29	32	32	32	32	32	32
Story Recall L2 Time 2	Correlation Coefficient	.190	-.110	.126	.321	.474(**)	.080	.049	.188	.054	1.000
	Sig. (2-tailed)	.306	.556	.509	.089	.006	.664	.790	.303	.771	.
	N	31	31	30	29	32	32	32	32	32	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D ii: Spearman correlations between WM and oral data at Time 1

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Question total	Correlation Coefficient	.316	-.125	-.052	-.495(**)	.006
	Sig. (2-tailed)	.083	.501	.784	.006	.974
	N	31	31	30	29	32
Question ratio	Correlation Coefficient	.139	-.100	.027	-.338	-.002
	Sig. (2-tailed)	.454	.592	.886	.073	.993
	N	31	31	30	29	32
Copula	Correlation Coefficient	.287	.031	-.006	-.542(**)	-.092
	Sig. (2-tailed)	.117	.869	.977	.002	.615
	N	31	31	30	29	32
Lexical	Correlation Coefficient	.285	-.259	-.170	-.131	.085
	Sig. (2-tailed)	.120	.160	.368	.498	.644
	N	31	31	30	29	32
Complex	Correlation Coefficient	.001	-.118	-.219	-.059	-.013
	Sig. (2-tailed)	.996	.528	.244	.759	.942
	N	31	31	30	29	32
Nontarget	Correlation Coefficient	.198	.048	.046	-.018	.288
	Sig. (2-tailed)	.286	.797	.810	.928	.110
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D iii: Spearman correlations between WM at Time 2 and oral data at Time 2

		Listening Span Time 2	Digits Back L1 Time 2	Digits Back L2 Time 2	Story Recall L1 Time 2	Story Recall L2 Time 2
Question total Time 2	Correlation Coefficient	-.162	-.310	-.010	-.065	.166
	Sig. (2-tailed)	.376	.084	.955	.725	.364
	N	32	32	32	32	32
Question ratio Time 2	Correlation Coefficient	-.225	-.258	.139	-.092	.160
	Sig. (2-tailed)	.216	.154	.448	.615	.383
	N	32	32	32	32	32
Copula Time 2	Correlation Coefficient	-.188	-.262	-.032	-.071	.094
	Sig. (2-tailed)	.303	.148	.861	.698	.607
	N	32	32	32	32	32
Lexical Time 2	Correlation Coefficient	-.123	-.151	-.042	.072	.107
	Sig. (2-tailed)	.502	.409	.821	.695	.558
	N	32	32	32	32	32
Complex Time 2	Correlation Coefficient	.029	-.291	.151	.030	.185
	Sig. (2-tailed)	.877	.106	.409	.872	.311
	N	32	32	32	32	32
Nontarget Time 2	Correlation Coefficient	-.110	-.069	-.283	-.215	-.201
	Sig. (2-tailed)	.548	.709	.116	.237	.269
	N	32	32	32	32	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D iv: Spearman correlations between WM at Time 1 and oral data at Time 2

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Question total Time 2	Correlation Coefficient	.236	-.320	.164	-.034	.136
	Sig. (2-tailed)	.202	.079	.385	.863	.459
	N	31	31	30	29	32
Question ratio Time 2	Correlation Coefficient	.129	-.454(*)	.185	.034	.021
	Sig. (2-tailed)	.489	.010	.327	.862	.908
	N	31	31	30	29	32
Copula Time 2	Correlation Coefficient	.303	-.220	.166	-.146	.074
	Sig. (2-tailed)	.097	.235	.381	.449	.686
	N	31	31	30	29	32
Lexical Time 2	Correlation Coefficient	.000	-.365(*)	-.013	.215	.087
	Sig. (2-tailed)	.998	.044	.945	.263	.635
	N	31	31	30	29	32
Complex Time 2	Correlation Coefficient	.114	-.243	-.066	.138	-.137
	Sig. (2-tailed)	.542	.188	.729	.476	.456
	N	31	31	30	29	32
Nontarget Time 2	Correlation Coefficient	-.121	.043	-.224	-.096	-.011
	Sig. (2-tailed)	.518	.817	.235	.619	.952
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D v: Spearman correlations between WM at Time 1 and change in oral data, Time 1 and Time 2

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Change OQT total	Correlation Coefficient	-.038	-.256	.150	.389(*)	.069
	Sig. (2-tailed)	.839	.164	.430	.037	.706
	N	31	31	30	29	32
Change OQT ratio	Correlation Coefficient	-.026	-.206	.217	.367	.030
	Sig. (2-tailed)	.889	.265	.249	.050	.872
	N	31	31	30	29	32
Change Copula	Correlation Coefficient	.056	-.187	.081	.271	.090
	Sig. (2-tailed)	.765	.313	.670	.155	.623
	N	31	31	30	29	32
Change Lexical	Correlation Coefficient	-.210	-.162	.158	.358	.111
	Sig. (2-tailed)	.257	.384	.405	.057	.547
	N	31	31	30	29	32
Change Complex	Correlation Coefficient	.082	.022	.108	.164	-.113
	Sig. (2-tailed)	.660	.907	.570	.394	.537
	N	31	31	30	29	32
Change Nontarget	Correlation Coefficient	-.245	-.036	-.298	-.061	-.260
	Sig. (2-tailed)	.183	.847	.110	.752	.151
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D vi: Spearman correlations between WM at Time 1 and RT Speeds at Time 1

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
RT total	Correlation Coefficient	.153	.152	.057	.439(*)	-.106
	Sig. (2-tailed)	.411	.415	.764	.017	.563
	N	31	31	30	29	32
RT grammatical	Correlation Coefficient	.153	.071	-.045	.440(*)	-.127
	Sig. (2-tailed)	.412	.704	.815	.017	.490
	N	31	31	30	29	32
RT ungrammatical	Correlation Coefficient	.149	.232	.134	.411(*)	-.111
	Sig. (2-tailed)	.423	.209	.480	.027	.546
	N	31	31	30	29	32
RT short movement	Correlation Coefficient	.137	.134	.005	.448(*)	-.057
	Sig. (2-tailed)	.463	.473	.978	.015	.757
	N	31	31	30	29	32
RT long movement	Correlation Coefficient	.171	.122	.016	.401(*)	-.139
	Sig. (2-tailed)	.357	.512	.935	.031	.448
	N	31	31	30	29	32
RT subjacency	Correlation Coefficient	.155	.228	.129	.345	-.140
	Sig. (2-tailed)	.405	.218	.497	.066	.443
	N	31	31	30	29	32
RT object questions	Correlation Coefficient	.143	.079	.093	.460(*)	-.050
	Sig. (2-tailed)	.443	.672	.626	.012	.786
	N	31	31	30	29	32
RT subject questions	Correlation Coefficient	.132	.202	-.003	.427(*)	-.118
	Sig. (2-tailed)	.478	.277	.987	.021	.519
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D vii: Spearman correlations between WM at Time 2 and RT Speeds at Time 2

		Listening Span Time 2	Digits Back L1 Time 2	Digits Back L2 Time 2	Story Recall L1 Time 2	Story Recall L2 Time 2
RT total Time 2	Correlation Coefficient	-.001	.007	-.058	.194	.528(**)
	Sig. (2-tailed)	.994	.968	.751	.288	.002
	N	32	32	32	32	32
RT grammatical Time 2	Correlation Coefficient	-.002	-.003	-.041	.254	.484(**)
	Sig. (2-tailed)	.990	.986	.825	.160	.005
	N	32	32	32	32	32
RT ungrammatical Time 2	Correlation Coefficient	-.021	.053	-.075	.184	.498(**)
	Sig. (2-tailed)	.910	.775	.685	.312	.004
	N	32	32	32	32	32
RT short movement Time 2	Correlation Coefficient	-.002	-.032	-.050	.171	.442(*)
	Sig. (2-tailed)	.993	.864	.786	.350	.011
	N	32	32	32	32	32
RT long movement Time 2	Correlation Coefficient	-.013	.030	-.074	.228	.495(**)
	Sig. (2-tailed)	.942	.873	.687	.209	.004
	N	32	32	32	32	32
RT subjacency Time 2	Correlation Coefficient	.034	.222	.074	.192	.590(**)
	Sig. (2-tailed)	.853	.221	.687	.291	.000
	N	32	32	32	32	32
RT object questions Time 2	Correlation Coefficient	-.047	-.026	-.078	.158	.459(**)
	Sig. (2-tailed)	.799	.886	.671	.389	.008
	N	32	32	32	32	32
RT subject questions Time 2	Correlation Coefficient	.024	.004	-.081	.190	.446(*)
	Sig. (2-tailed)	.896	.985	.659	.298	.011
	N	32	32	32	32	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D viii: Spearman correlations between WM at Time 1 and RT speeds at Time 2

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
RT total Time 2	Correlation Coefficient	-.016	.028	-.026	.581(**)	.389(*)
	Sig. (2-tailed)	.933	.882	.893	.001	.028
	N	31	31	30	29	32
RT grammatical Time 2	Correlation Coefficient	-.042	.052	.004	.624(**)	.389(*)
	Sig. (2-tailed)	.822	.780	.985	.000	.028
	N	31	31	30	29	32
RT ungrammatical Time 2	Correlation Coefficient	-.014	.047	-.028	.514(**)	.375(*)
	Sig. (2-tailed)	.939	.803	.885	.004	.034
	N	31	31	30	29	32
RT short movement Time 2	Correlation Coefficient	-.097	-.024	-.114	.608(**)	.327
	Sig. (2-tailed)	.604	.899	.550	.000	.068
	N	31	31	30	29	32
RT long movement Time 2	Correlation Coefficient	-.018	.105	.024	.580(**)	.420(*)
	Sig. (2-tailed)	.924	.575	.901	.001	.017
	N	31	31	30	29	32
RT subjacency Time 2	Correlation Coefficient	.060	.122	.108	.490(**)	.324
	Sig. (2-tailed)	.749	.515	.569	.007	.071
	N	31	31	30	29	32
RT object questions Time 2	Correlation Coefficient	-.010	.032	-.042	.574(**)	.304
	Sig. (2-tailed)	.957	.866	.827	.001	.090
	N	31	31	30	29	32
RT subject questions Time 2	Correlation Coefficient	-.059	.040	-.052	.631(**)	.362(*)
	Sig. (2-tailed)	.754	.831	.785	.000	.041
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D ix: Spearman correlations between WM at Time 1 and change in RT speeds between Time 1 and Time 2

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Change total RT	Correlation Coefficient	-.251	-.054	-.064	-.001	.402(*)
	Sig. (2-tailed)	.173	.774	.736	.997	.022
	N	31	31	30	29	32
Change grammatical RT	Correlation Coefficient	-.227	-.005	-.026	.032	.346
	Sig. (2-tailed)	.219	.979	.889	.870	.052
	N	31	31	30	29	32
Change ungrammatical RT	Correlation Coefficient	-.201	-.166	-.100	-.002	.388(*)
	Sig. (2-tailed)	.278	.372	.600	.991	.028
	N	31	31	30	29	32
Change short movement RT	Correlation Coefficient	-.264	.000	-.019	.076	.300
	Sig. (2-tailed)	.151	1.000	.922	.696	.096
	N	31	31	30	29	32
Change long movement RT	Correlation Coefficient	-.217	-.073	-.043	-.003	.382(*)
	Sig. (2-tailed)	.242	.694	.821	.987	.031
	N	31	31	30	29	32
Change subjacency RT	Correlation Coefficient	-.088	-.096	-.004	.029	.409(*)
	Sig. (2-tailed)	.637	.608	.981	.883	.020
	N	31	31	30	29	32
Change object RT	Correlation Coefficient	-.242	-.008	-.071	-.066	.220
	Sig. (2-tailed)	.189	.967	.711	.736	.226
	N	31	31	30	29	32
Change subject RT	Correlation Coefficient	-.252	-.112	-.026	.063	.425(*)
	Sig. (2-tailed)	.172	.548	.890	.745	.015
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table D x: Spearman correlations between WM at Time 1 and RT Accuracy scores at Time 1

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Accuracy	Correlation Coefficient	.022	.094	-.176	.192	-.032
	Sig. (2-tailed)	.906	.613	.352	.319	.861
	N	31	31	30	29	32
Grammatical	Correlation Coefficient	.013	-.097	-.115	-.182	.021
	Sig. (2-tailed)	.944	.603	.546	.345	.911
	N	31	31	30	29	32
Ungrammatical	Correlation Coefficient	.040	.246	-.093	.368(*)	-.114
	Sig. (2-tailed)	.832	.182	.623	.049	.536
	N	31	31	30	29	32
Short movement	Correlation Coefficient	-.021	.119	-.157	.247	-.050
	Sig. (2-tailed)	.909	.525	.408	.196	.785
	N	31	31	30	29	32
Long movement	Correlation Coefficient	.164	-.131	-.224	-.439(*)	-.155
	Sig. (2-tailed)	.377	.482	.234	.017	.396
	N	31	31	30	29	32
Subjacency	Correlation Coefficient	.145	.271	-.006	.402(*)	.014
	Sig. (2-tailed)	.435	.140	.974	.030	.940
	N	31	31	30	29	32
Object questions	Correlation Coefficient	.174	-.027	-.210	.093	-.196
	Sig. (2-tailed)	.350	.885	.266	.633	.283
	N	31	31	30	29	32
Subject questions	Correlation Coefficient	-.088	.003	-.219	-.160	-.085
	Sig. (2-tailed)	.639	.987	.246	.406	.645
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D xi: Spearman correlations between WM at Time 2 and RT accuracy scores at Time 2

		Listening Span Time 2	Digits Back L1 Time 2	Digits Back L2 Time 2	Story Recall L1 Time 2	Story Recall L2 Time 2
Accuracy Time 2	Correlation Coefficient	.119	-.062	.209	-.035	.243
	Sig. (2-tailed)	.516	.734	.252	.848	.179
	N	32	32	32	32	32
Gram accuracy Time 2	Correlation Coefficient	.033	-.132	.188	-.112	.036
	Sig. (2-tailed)	.857	.473	.304	.542	.845
	N	32	32	32	32	32
Ungram accuracy Time 2	Correlation Coefficient	.040	-.057	.066	.053	.198
	Sig. (2-tailed)	.827	.757	.718	.772	.278
	N	32	32	32	32	32
Short movement Time 2	Correlation Coefficient	.054	-.004	.345	.039	.115
	Sig. (2-tailed)	.768	.981	.053	.832	.531
	N	32	32	32	32	32
Long movement Time 2	Correlation Coefficient	.080	-.202	.028	-.327	-.021
	Sig. (2-tailed)	.662	.266	.878	.068	.910
	N	32	32	32	32	32
Subjacency Time 2	Correlation Coefficient	.171	.038	-.054	.180	.275
	Sig. (2-tailed)	.349	.837	.767	.324	.128
	N	32	32	32	32	32
Object questions Time 2	Correlation Coefficient	.206	.105	.228	-.153	.010
	Sig. (2-tailed)	.258	.569	.210	.404	.955
	N	32	32	32	32	32
Subject questions Time 2	Correlation Coefficient	-.126	-.268	.168	-.119	.093
	Sig. (2-tailed)	.491	.138	.358	.516	.612
	N	32	32	32	32	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D xii: Spearman correlations between WM at Time 1 and RT Accuracy scores at Time 2

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Accuracy Time 2	Correlation Coefficient	-.023	-.105	.254	-.043	.017
	Sig. (2-tailed)	.903	.573	.176	.824	.926
	N	31	31	30	29	32
Grammatical accuracy Time 2	Correlation Coefficient	.102	-.287	-.029	-.470(*)	-.115
	Sig. (2-tailed)	.584	.117	.878	.010	.530
	N	31	31	30	29	32
Ungrammatical accuracy Time 2	Correlation Coefficient	-.094	.097	.321	.352	.072
	Sig. (2-tailed)	.617	.605	.083	.061	.697
	N	31	31	30	29	32
Short movement Time 2	Correlation Coefficient	.019	-.073	.363(*)	.116	-.145
	Sig. (2-tailed)	.920	.696	.049	.548	.430
	N	31	31	30	29	32
Long movement Time 2	Correlation Coefficient	-.056	-.255	-.147	-.437(*)	-.128
	Sig. (2-tailed)	.763	.167	.438	.018	.484
	N	31	31	30	29	32
Subjacency Time 2	Correlation Coefficient	-.102	.126	.216	.115	.277
	Sig. (2-tailed)	.587	.498	.252	.552	.124
	N	31	31	30	29	32
Object questions Time 2	Correlation Coefficient	-.124	.016	.267	-.046	-.033
	Sig. (2-tailed)	.505	.930	.154	.813	.859
	N	31	31	30	29	32
Subject questions Time 2	Correlation Coefficient	.052	-.363(*)	.050	-.236	-.197
	Sig. (2-tailed)	.780	.045	.795	.217	.280
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D xiii: Spearman correlations between WM and changes in Accuracy scores between Time 1 and Time 2

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Change Accuracy	Correlation Coefficient	.016	-.127	.348	-.181	.038
	Sig. (2-tailed)	.930	.496	.060	.346	.837
	N	31	31	30	29	32
Change grammatical accuracy	Correlation Coefficient	.073	-.130	.084	-.255	-.197
	Sig. (2-tailed)	.696	.486	.660	.182	.281
	N	31	31	30	29	32
Change ungrammatical accuracy	Correlation Coefficient	-.109	-.190	.295	-.056	.230
	Sig. (2-tailed)	.558	.307	.113	.772	.206
	N	31	31	30	29	32
Change short movement accuracy	Correlation Coefficient	.131	-.195	.334	-.148	-.009
	Sig. (2-tailed)	.481	.293	.071	.444	.959
	N	31	31	30	29	32
Change long movement accuracy	Correlation Coefficient	-.145	-.065	.065	.060	-.026
	Sig. (2-tailed)	.436	.728	.734	.757	.886
	N	31	31	30	29	32
Change subjacency accuracy	Correlation Coefficient	-.157	-.129	.172	-.182	.200
	Sig. (2-tailed)	.398	.491	.364	.345	.271
	N	31	31	30	29	32
Change object accuracy	Correlation Coefficient	-.198	-.046	.252	-.132	.079
	Sig. (2-tailed)	.286	.805	.180	.493	.666
	N	31	31	30	29	32
Change subject accuracy	Correlation Coefficient	.218	-.125	.286	-.095	-.119
	Sig. (2-tailed)	.239	.502	.125	.625	.518
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table D xiv: Kruskal-Wallis Test Statistics for between-group differences in WM at Time 1 (grouped by Accuracy at Time 2)

	Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Chi-Square	.863	2.063	2.160	.198	.799
Df	2	2	2	2	2
Asymp. Sig.	.650	.356	.340	.906	.671

Table D xv: Kruskal-Wallis Test Statistics for between-group differences in changes in RT accuracy between Time 1 and Time 2 (grouped by Accuracy at Time 2)

(a) for total, grammatical and ungrammatical accuracy

	Change Accuracy	Change grammatical accuracy	Change ungrammatical accuracy
Chi-Square	3.830	4.717	1.056
Df	2	2	2
Asymp. Sig.	.147	.095	.590

(b) for submeasures

	Change short movement accuracy	Change long movement accuracy	Change subjacency accuracy	Change object accuracy	Change subject accuracy
Chi-Square	2.029	3.314	.903	3.727	2.266
Df	2	2	2	2	2
Asymp. Sig.	.363	.191	.637	.155	.322

Table D xvi: Spearman correlations between WM and RT Accuracy at Time 2 and Change in RT Accuracy, split by group (grouped according to Accuracy scores at Time 2)
(a) HIGH group, WM and Accuracy Time 2

			Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
High	Accuracy Time 2	Correlation Coefficient	.618(*)	.234	.519	.279	-.097
		Sig. (2-tailed)	.025	.441	.084	.356	.742
		N	13	13	12	13	14
	Gram accuracy Time 2	Correlation Coefficient	.109	-.277	.124	-.596(*)	-.550(*)
		Sig. (2-tailed)	.722	.359	.702	.032	.041
		N	13	13	12	13	14
	Ungram accuracy Time 2	Correlation Coefficient	.345	.364	.274	.692(**)	.347
		Sig. (2-tailed)	.248	.222	.389	.009	.224
		N	13	13	12	13	14
	Short movement Time 2	Correlation Coefficient	.624(*)	-.129	.522	.357	-.156
		Sig. (2-tailed)	.023	.674	.082	.231	.594
		N	13	13	12	13	14
	Long movement Time 2	Correlation Coefficient	-.227	-.100	-.191	-.670(*)	-.516
		Sig. (2-tailed)	.455	.746	.552	.012	.059
		N	13	13	12	13	14
	Subjacency Time 2	Correlation Coefficient	.300	.356	.462	.501	.521
		Sig. (2-tailed)	.320	.233	.130	.081	.056
		N	13	13	12	13	14
	Object questions Time 2	Correlation Coefficient	.046	.110	.123	-.034	-.220
		Sig. (2-tailed)	.880	.721	.704	.913	.450
		N	13	13	12	13	14
	Subject questions Time 2	Correlation Coefficient	.462	-.377	.274	-.221	-.483
		Sig. (2-tailed)	.112	.204	.390	.468	.080
		N	13	13	12	13	14

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table D xvi: Spearman correlations between WM and RT Accuracy at Time 2 and Change in RT Accuracy, split by group (grouped according to Accuracy scores at Time 2)
(b) HIGH group – WM and Change in RT Accuracy

			Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
High	Change Accuracy	Correlation Coefficient	.006	-.244	.385	-.105	.164
		Sig. (2-tailed)	.986	.422	.217	.734	.575
		N	13	13	12	13	14
	Change gram accuracy	Correlation Coefficient	-.003	-.387	-.025	.010	-.029
		Sig. (2-tailed)	.993	.192	.939	.975	.922
		N	13	13	12	13	14
	Change ungram accuracy	Correlation Coefficient	.081	-.036	.439	-.003	.330
		Sig. (2-tailed)	.791	.907	.154	.993	.250
		N	13	13	12	13	14
	Change Short movement accuracy	Correlation Coefficient	.235	-.392	.459	.048	.123
		Sig. (2-tailed)	.439	.186	.134	.875	.676
		N	13	13	12	13	14
	Change Long movement accuracy	Correlation Coefficient	-.054	-.366	-.328	.243	.044
		Sig. (2-tailed)	.862	.219	.298	.424	.882
		N	13	13	12	13	14
	Change Subjacency accuracy	Correlation Coefficient	-.107	-.001	.287	-.436	.131
		Sig. (2-tailed)	.728	.996	.365	.137	.655
		N	13	13	12	13	14
	Change Object accuracy	Correlation Coefficient	-.389	.043	.271	-.054	.397
		Sig. (2-tailed)	.189	.889	.394	.861	.160
		N	13	13	12	13	14
	Change Subject accuracy	Correlation Coefficient	.346	-.504	.187	.131	-.088
		Sig. (2-tailed)	.247	.079	.561	.669	.764
		N	13	13	12	13	14

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table D xvi: Spearman correlations between WM and RT Accuracy at Time 2 and Change in RT Accuracy, split by group (grouped according to Accuracy scores at Time 2)
(c) MID group – WM and Accuracy at Time 2

			Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Mid	Accuracy Time 2	Correlation Coefficient	-.058	.287	.143	-.167	-.141
		Sig. (2-tailed)	.866	.392	.675	.645	.679
		N	11	11	11	10	11
	Gram accuracy Time 2	Correlation Coefficient	.563	-.284	-.315	-.726(*)	-.096
		Sig. (2-tailed)	.071	.397	.345	.018	.779
		N	11	11	11	10	11
	Ungram accuracy Time 2	Correlation Coefficient	-.469	.359	.312	.546	-.152
		Sig. (2-tailed)	.146	.278	.350	.102	.656
		N	11	11	11	10	11
	Short movement Time 2	Correlation Coefficient	-.037	.274	.148	.143	-.608(*)
		Sig. (2-tailed)	.914	.414	.664	.694	.047
		N	11	11	11	10	11
	Long movement Time 2	Correlation Coefficient	.276	-.359	-.427	-.489	.080
		Sig. (2-tailed)	.412	.278	.191	.151	.814
		N	11	11	11	10	11
	Subjacency Time 2	Correlation Coefficient	-.527	.089	.014	.245	.309
		Sig. (2-tailed)	.096	.794	.967	.496	.355
		N	11	11	11	10	11
	Object questions Time 2	Correlation Coefficient	.090	.447	.376	.006	-.092
		Sig. (2-tailed)	.793	.168	.254	.987	.788
		N	11	11	11	10	11
	Subject questions Time 2	Correlation Coefficient	.048	-.515	-.644(*)	-.234	-.373
		Sig. (2-tailed)	.889	.105	.033	.514	.259
		N	11	11	11	10	11

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table D xvi: Spearman correlations between WM and RT Accuracy at Time 2 and Change in RT Accuracy, split by group (grouped according to Accuracy scores at Time 2)
(d) MID group – WM and Change in Accuracy

			Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Mid	Change Accuracy	Correlation Coefficient	.178	.309	.360	-.091	.201
		Sig. (2-tailed)	.600	.355	.277	.803	.554
		N	11	11	11	10	11
	Change gram accuracy	Correlation Coefficient	.253	.327	.009	-.362	-.368
		Sig. (2-tailed)	.453	.326	.979	.304	.266
		N	11	11	11	10	11
	Change ungram accuracy	Correlation Coefficient	.057	-.098	.105	.091	.447
		Sig. (2-tailed)	.868	.773	.759	.803	.168
		N	11	11	11	10	11
	Change Short movement accuracy	Correlation Coefficient	.332	.182	.198	-.189	.124
		Sig. (2-tailed)	.319	.591	.560	.601	.715
		N	11	11	11	10	11
	Change Long movement accuracy	Correlation Coefficient	-.037	.382	.209	-.137	-.149
		Sig. (2-tailed)	.913	.246	.538	.706	.662
		N	11	11	11	10	11
	Change Subjacency accuracy	Correlation Coefficient	-.072	-.313	-.069	.128	.437
		Sig. (2-tailed)	.834	.349	.840	.724	.179
		N	11	11	11	10	11
	Change Object accuracy	Correlation Coefficient	.092	.406	.322	-.135	.127
		Sig. (2-tailed)	.788	.215	.334	.709	.710
		N	11	11	11	10	11
	Change Subject accuracy	Correlation Coefficient	.162	.398	.301	-.128	-.027
		Sig. (2-tailed)	.634	.225	.369	.725	.936
		N	11	11	11	10	11

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table D xvi: Spearman correlations between WM and RT Accuracy at Time 2 and Change in RT Accuracy, split by group (grouped according to Accuracy scores at Time 2)
(e) LOW group – WM and RT Accuracy at Time 2

			Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Low	Accuracy Time 2	Correlation Coefficient	.330	-.273	-.018	-.294	.073
		Sig. (2-tailed)	.469	.554	.969	.571	.877
		N	7	7	7	6	7
	Gram accuracy Time 2	Correlation Coefficient	.414	-.143	-.143	-.257	.357
		Sig. (2-tailed)	.355	.760	.760	.623	.432
		N	7	7	7	6	7
	Ungram accuracy Time 2	Correlation Coefficient	-.505	-.179	.036	-.657	-.607
		Sig. (2-tailed)	.248	.702	.939	.156	.148
		N	7	7	7	6	7
	Short movement Time 2	Correlation Coefficient	.136	-.234	.126	-.314	.180
		Sig. (2-tailed)	.771	.613	.788	.544	.699
		N	7	7	7	6	7
	Long movement Time 2	Correlation Coefficient	.464	-.414	-.414	.086	.288
		Sig. (2-tailed)	.295	.355	.355	.872	.531
		N	7	7	7	6	7
	Subjacency Time 2	Correlation Coefficient	-.148	-.055	-.128	-.912(*)	-.679
		Sig. (2-tailed)	.751	.907	.784	.011	.093
		N	7	7	7	6	7
	Object questions Time 2	Correlation Coefficient	.028	-.312	-.055	.177	.367
		Sig. (2-tailed)	.953	.496	.907	.738	.418
		N	7	7	7	6	7
	Subject questions Time 2	Correlation Coefficient	.273	-.270	-.054	-.371	.234
		Sig. (2-tailed)	.554	.558	.908	.468	.613
		N	7	7	7	6	7

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table D xvi: Spearman correlations between WM and RT Accuracy at Time 2 and Change in RT Accuracy, split by group (grouped according to Accuracy scores at Time 2)
(f) LOW group – WM and Change in Accuracy

			Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Low	Change Accuracy	Correlation Coefficient	-.396	-.357	.071	-.371	-.250
		Sig. (2-tailed)	.379	.432	.879	.468	.589
		N	7	7	7	6	7
	Change gram accuracy	Correlation Coefficient	-.182	-.216	.036	-.232	-.144
		Sig. (2-tailed)	.696	.641	.939	.658	.758
		N	7	7	7	6	7
	Change ungram accuracy	Correlation Coefficient	-.734	-.109	.364	-.145	-.200
		Sig. (2-tailed)	.060	.816	.423	.784	.667
		N	7	7	7	6	7
	Change Short movement accuracy	Correlation Coefficient	-.382	-.378	.054	-.493	-.360
		Sig. (2-tailed)	.398	.403	.908	.321	.427
		N	7	7	7	6	7
	Change Long movement accuracy	Correlation Coefficient	-.468	-.055	.273	.667	.000
		Sig. (2-tailed)	.290	.908	.554	.148	1.000
		N	7	7	7	6	7
	Change Subjacency accuracy	Correlation Coefficient	-.618	.270	.559	.377	.000
		Sig. (2-tailed)	.139	.558	.192	.461	1.000
		N	7	7	7	6	7
	Change Object accuracy	Correlation Coefficient	-.450	-.571	-.143	-.257	-.286
		Sig. (2-tailed)	.310	.180	.760	.623	.535
		N	7	7	7	6	7
	Change Subject accuracy	Correlation Coefficient	-.306	-.321	.179	-.657	-.357
		Sig. (2-tailed)	.504	.482	.702	.156	.432
		N	7	7	7	6	7

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table D xvii: Kruskal-Wallis Test Statistics for between-group differences on Accuracy at Time 2 and Change in Accuracy (grouped by WM test)

(a) Listening Span

	Accuracy Time 2	Change Accuracy
Chi-Square	1.513	.261
Df	2	2
Asymp. Sig.	.469	.878

(b) Digits Back L1

	Accuracy Time 2	Change Accuracy
Chi-Square	1.656	3.768
Df	2	2
Asymp. Sig.	.437	.152

(c) Digits Back L2

	Accuracy Time 2	Change Accuracy
Chi-Square	2.899	4.055
Df	2	2
Asymp. Sig.	.235	.132

(d) Story Recall L1

	Accuracy Time 2	Change Accuracy
Chi-Square	**8.792**	.543
Df	**2**	2
Asymp. Sig.	**.012**	.762

(e) Story Recall L2

	Accuracy Time 2	Change Accuracy
Chi-Square	.648	.249
Df	2	2
Asymp. Sig.	.723	.883

Table D xviii: Comparison of between-group means for Accuracy at Time 2 (grouped by Story Recall in L1)

Story Recall L1		N	Mean
Low	Accuracy Time 2	10	43.20
Mid	Accuracy Time 2	10	33.20
High	Accuracy Time 2	9	43.22

Table D xix: Kruskal Wallis Test Statistics of between-group differences on RT accuracy scores at Time 1 and Time 2 comparing groups split by Story Recall in L1 scores

(a) Accuracy scores at Time 1

	Accuracy	Gram	**Ungram**	Short movement	Long movement	**Subjacency**	Object questions	Subject questions
Chi-Square	3.225	1.257	**7.753**	2.326	5.874	**7.337**	.768	2.061
Df	2	2	**2**	2	2	**2**	2	2
Asymp. Sig.	.199	.533	**.021**	.313	.053	**.026**	.681	.357

(b) Accuracy scores at Time 2

	Accuracy Time 2	Gram accuracy Time 2	Ungram accuracy Time 2	Short movement Time 2	Long movement Time 2	Subjacency Time 2	Object questions Time 2	Subject questions Time 2
Chi-Square	8.792	5.943	9.814	8.444	6.093	4.574	9.030	4.040
Df	2	2	2	2	2	2	2	2
Asymp. Sig.	.012	.051	.007	.015	.048	.102	.011	.133

Table D xx: Comparison of means for subjacency accuracy scores at Time 1 and Time 2, split by Story Recall in L1

Story Recall L1 group		Mean
Low	Subjacency	5.10
	Subjacency Time 2	4.30
Mid	Subjacency	3.80
	Subjacency Time 2	3.20
High	Subjacency	7.78
	Subjacency Time 2	5.67

Table D xxi: Spearman correlations between WM and Accuracy at Time 2 (split by Story Recall in L1 scores)

Story Recall L1 Group			Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Low	Accuracy Time 2	Correlation Coefficient	-.243	-.227	.210	-.067	-.055
		Sig. (2-tailed)	.529	.528	.587	.854	.881
		N	9	10	9	10	10
	Gram accuracy Time 2	Correlation Coefficient	-.458	-.553	.026	-.178	.252
		Sig. (2-tailed)	.215	.097	.948	.622	.483
		N	9	10	9	10	10
	Ungram accuracy Time 2	Correlation Coefficient	.241	.285	.386	.150	-.245
		Sig. (2-tailed)	.533	.425	.305	.679	.496
		N	9	10	9	10	10
	Short movement Time 2	Correlation Coefficient	-.034	-.228	.369	.214	-.348
		Sig. (2-tailed)	.932	.526	.329	.553	.325
		N	9	10	9	10	10
	Long movement Time 2	Correlation Coefficient	-.517	-.413	-.142	-.387	.043
		Sig. (2-tailed)	.154	.236	.715	.269	.906
		N	9	10	9	10	10
	Subjacency Time 2	Correlation Coefficient	.180	.556	.198	.248	.019
		Sig. (2-tailed)	.644	.095	.609	.490	.959
		N	9	10	9	10	10
	Object questions Time 2	Correlation Coefficient	-.224	-.327	.297	-.012	.091
		Sig. (2-tailed)	.563	.356	.438	.973	.802
		N	9	10	9	10	10
	Subject questions Time 2	Correlation Coefficient	-.266	-.343	.013	.046	-.074
		Sig. (2-tailed)	.489	.332	.974	.899	.839
		N	9	10	9	10	10
Mid	Accuracy Time 2	Correlation Coefficient	-.228	-.402	.284	-.457	.311
		Sig. (2-tailed)	.526	.249	.427	.184	.382
		N	10	10	10	10	10
	Gram accuracy Time 2	Correlation Coefficient	-.209	-.322	.152	-.286	.492
		Sig. (2-tailed)	.562	.364	.675	.424	.148
		N	10	10	10	10	10
	Ungram accuracy Time 2	Correlation Coefficient	-.700(*)	-.254	.358	-.483	-.205
		Sig. (2-tailed)	.024	.479	.310	.157	.570
		N	10	10	10	10	10
	Short movement Time 2	Correlation Coefficient	-.143	-.332	.351	-.455	.209
		Sig. (2-tailed)	.693	.348	.320	.186	.562
		N	10	10	10	10	10

(cont'd)

Story Recall L1 Group Correlations (cont'd)			Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Mid	Long movement Time 2	Correlation Coefficient	-.149	-.330	.092	-.098	.428
		Sig. (2-tailed)	.682	.351	.801	.788	.217
		N	10	10	10	10	10
	Subjacency Time 2	Correlation Coefficient	-.687(*)	-.350	.009	-.625	.133
		Sig. (2-tailed)	.028	.322	.980	.053	.714
		N	10	10	10	10	10
	Object questions Time 2	Correlation Coefficient	-.022	-.192	.230	-.126	.318
		Sig. (2-tailed)	.951	.595	.523	.729	.371
		N	10	10	10	10	10
	Subject questions Time 2	Correlation Coefficient	-.220	-.346	.321	-.474	.291
		Sig. (2-tailed)	.542	.328	.366	.166	.415
		N	10	10	10	10	10
High	Accuracy Time 2	Correlation Coefficient	.812(**)	-.458	-.072	.618	.336
		Sig. (2-tailed)	.008	.215	.866	.076	.376
		N	9	9	8	9	9
	Gram accuracy Time 2	Correlation Coefficient	.286	-.485	-.518	-.253	-.473
		Sig. (2-tailed)	.456	.185	.188	.511	.199
		N	9	9	8	9	9
	Ungram accuracy Time 2	Correlation Coefficient	.433	-.134	-.143	.502	.653
		Sig. (2-tailed)	.244	.731	.736	.168	.057
		N	9	9	8	9	9
	Short movement Time 2	Correlation Coefficient	.862(**)	-.382	.120	.723(*)	-.139
		Sig. (2-tailed)	.003	.310	.778	.028	.722
		N	9	9	8	9	9
	Long movement Time 2	Correlation Coefficient	-.067	-.662	-.771(*)	-.439	-.198
		Sig. (2-tailed)	.864	.052	.025	.237	.609
		N	9	9	8	9	9
	Subjacency Time 2	Correlation Coefficient	.203	-.004	-.048	.364	.699(*)
		Sig. (2-tailed)	.601	.991	.910	.335	.036
		N	9	9	8	9	9
	Object questions Time 2	Correlation Coefficient	.608	-.140	.145	.538	.004
		Sig. (2-tailed)	.083	.720	.733	.135	.991
		N	9	9	8	9	9
	Subject questions Time 2	Correlation Coefficient	.332	-.898(**)	-.822(*)	.056	-.312
		Sig. (2-tailed)	.383	.001	.012	.887	.414
		N	9	9	8	9	9

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table D xxii: Kruskal Wallis Test Statistics of between-group differences on RT speeds at Time 1 and Time 2 comparing groups split by Story Recall in L1 scores
(a) RT speeds at Time 1

	RT	RT gram	RT ungram	RT Short movement	RT Long movement	**RT Subjacency**	RT Object questions	RT Subject questions
Chi-Square	9.083	8.939	8.316	8.621	7.419	**6.673**	9.754	7.651
Df	2	2	2	2	2	**2**	2	2
Asymp. Sig.	.011	.011	.016	.013	.024	**.036**	.008	.022

(b) RT speeds at Time 2

	RT Time 2	RT gram Time 2	RT ungram Time 2	RT Short movement Time 2	RT Long movement Time 2	RT Subjacency Time 2	RT Object questions Time 2	RT Subject questions Time 2
Chi-Square	7.997	9.979	5.698	9.544	8.185	4.461	8.471	9.132
Df	2	2	2	2	2	2	2	2
Asymp. Sig.	.018	.007	.058	.008	.017	.107	.014	.010

Table D xxiii: Comparison of means for subjacency RT speeds at Time 1 and Time 2 (in seconds), split by Story Recall in L1

Story Recall L1 group		Mean
Low	RT Subjacency	86.95
	RT Subjacency Time 2	66.98
Mid	RT Subjacency	76.27
	RT Subjacency Time 2	73.61
High	RT Subjacency	118.14
	RT Subjacency Time 2	102.33

Table D xxiv: Spearman correlations between oral fluency data and WM measures

		Listening Span	Digits Back L1	Digits Back L2	Story Recall L1	Story Recall L2
Type token ratio	Correlation Coefficient	-.123	.132	.292	.055	.245
	Sig. (2-tailed)	.511	.479	.117	.775	.176
	N	31	31	30	29	32
Type token ratio Time 2	Correlation Coefficient	.023	.177	.419(*)	-.130	-.053
	Sig. (2-tailed)	.901	.340	.021	.503	.771
	N	31	31	30	29	32
Repairs and filled pauses	Correlation Coefficient	.161	-.194	-.147	.046	-.065
	Sig. (2-tailed)	.387	.295	.437	.812	.723
	N	31	31	30	29	32
Repairs and filled pauses Time 2	Correlation Coefficient	-.073	-.136	-.250	.257	.234
	Sig. (2-tailed)	.694	.467	.182	.178	.197
	N	31	31	30	29	32

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).